National Library of Canada
Acquisitions and Bibliographic Services Branch
395 Wellington Street
Ottawa, Ontario
K1A 0N4

Bibliothèque nationale du Canada
Direction des acquisitions et des services bibliographiques
395, rue Wellington
Ottawa (Ontario)
K1A 0N4

Your file Votre référence

Our file Notre référence

NOTICE

The quality of this microform is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

Reproduction in full or in part of this microform is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30, and subsequent amendments.

AVIS

La qualité de cette microforme dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

La reproduction, même partielle, de cette microforme est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30, et ses amendements subséquents.

Canada

UNIVERSITY OF ALBERTA

The World, Others and the Self:
Philosophy and Its Epistemic Neuroses

by

Michael Hymers

©

A thesis submitted to the Faculty of Graduate Studies and Research
in partial fulfilment of the requirements for the degree of
Doctor of Philosophy.

Department of Philosophy

Edmonton, Alberta
Fall, 1993

National Library of Canada
Acquisitions and Bibliographic Services Branch
395 Wellington Street
Ottawa, Ontario
K1A 0N4

Bibliothèque nationale du Canada
Direction des acquisitions et des services bibliographiques
395, rue Wellington
Ottawa (Ontario)
K1A 0N4

Your file Votre référence

Our file Notre référence

The author has granted an irrevocable non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of his/her thesis by any means and in any form or format, making this thesis available to interested persons.

L'auteur a accordé une licence irrévocable et non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de sa thèse de quelque manière et sous quelque forme que ce soit pour mettre des exemplaires de cette thèse à la disposition des personnes intéressées.

The author retains ownership of the copyright in his/her thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without his/her permission.

L'auteur conserve la propriété du droit d'auteur qui protège sa thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

ISBN 0-315-88195-X

Canada

UNIVERSITY OF ALBERTA

RELEASE FORM

NAME OF AUTHOR: Michael Hymers

TITLE OF THESIS: The World, Others and the Self: Philosophy and Its Epistemic Neuroses

DEGREE: Doctor of Philosophy

YEAR THIS DEGREE GRANTED: 1993

Permission is hereby granted to the University of Alberta Library to reproduce single copies of this thesis and to lend or sell such copies for private, scholarly or scientific research-purposes only.

The author reserves all other publication and other rights in association with the copyright in the thesis, and except as hereinbefore provided neither the thesis nor any substantial portion thereof may be printed or otherwise reproduced in any material form whatever without the author's prior written permission.

Michael Hymers

9624 - 105 St., #607
Edmonton, Alberta
Canada T5K 0Z8

October 13, 1993

UNIVERSITY OF ALBERTA

FACULTY OF GRADUATE STUDIES AND RESEARCH

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research for acceptance, a thesis entitled "The World, Others and the Self: Philosophy and Its Epistemic Neuroses" submitted by Michael Hymers in partial fulfilment of the requirements for the degree of Doctor of Philosophy.

Dr. Bruce A. Hunter

Dr. Peter A. Schouls

Dr. Wesley E. Cooper

Dr. Dianne Chisholm

Dr. Richard Rorty

October 8, 1993

For Kathryn

ABSTRACT

I explore dichotomous treatments of subjectivity and objectivity in accounts of knowledge of the world, others and the self. In Chapters 1-5 I argue that realism and relativism suffer forms of *epistemic neurosis*--each is undermined by its own account of objectivity.

The realist sees the world and semantic notions, like truth and reference, as independent of our abilities to know about them. This implies the self-defeating result that we could be totally wrong about the meanings of our words, since meaning depends on our interaction with the world and with each other.

The relativist sees rationality, truth and concepts as local to cultures, eras or persons. But this view defeats itself too, since any argument for it must use a non-relativist account of truth and rationality. It is tacitly realistic, treating other cultures and minds as independent of our epistemic capacities. Thinkers who accept the objective-subjective dualism can be driven to relativism by a distaste for scientism and a failure to see the contingent, historical character of the alliance of realism and scientism, which survives in varieties of Marxism and analytical philosophy.

I avoid these problems by taking objectivity to imply that the world, other cultures and other minds are independent, not of what we *could* know, but only of our actual beliefs about them. Justification, truth and reference are normative--not objective, natural phenomena in the realist's sense, but not merely relative either.

In Chapter 6 I consider the status of self-knowledge in the work of Descartes, Hume and Kant, and in Chapter 7 I examine recent efforts to explain the asymmetry in the justification of first- and other-person knowledge-claims. Linguistic competence, I argue, entails being able to articulate one's intentional attitudes, but not (with the same reliability) those of others. But this asymmetry is not absolute; it varies with the degree of experience and justifiable trust shared by particular persons. Finally, in Chapter 8 I argue that worries raised in radical theory about "unified subjects" do not rule out seeing the self as unified in its recursive capacity for self-description.

ACKNOWLEDGEMENTS

The nature of intellectual influence and its appropriate acknowledgement might itself stand lengthy analysis, but I forego such theoretical niceties, since there are a few names I would like to mention, if only to present a certain image of myself as I would have others understand the author of this work.

First, if the phrase "Thanks to Bruce Hunter" seems repeated with near-comic frequency in the footnotes below, it remains the case that my greatest immediate debt is to Bruce, under whose supervision I have had the pleasure of working. His eye for subtle detail, his ability to understand what I have been doing (often before I do) and his respect for my work and my interests have been invaluable to me. I must also thank Wes Cooper for his encouragement and kind words and Peter Schouls for his astute advice in philosophical and other matters.

The roots of my interest in the issues broached here I can trace--at least in retrospect--to time spent at Dalhousie University in Halifax, where Steven Burns introduced me to the work of Marx and Wittgenstein. I have Tom Vinci to thank for my interest in Kant, Duncan MacIntosh for putting the words 'realism' and 'anti-realism' into my vocabulary and Richmond Campbell and Susan Sherwin for inspiring me to think about the relation between realism and radical theory. Outside the Philosophy Department I need also to thank my good friend Charles Davidson for convincing me, in ways that he may not even realize, that the issues taken up below are worth taking up.

At the University of Alberta I have benefited from discussion and association with my friends and colleagues in the graduate-programme. I mention especially Tilman Lichter, Edrie Sobstyl, Joachim Ludwig, Guangwei Ouyang, Susan Turner and Stephen Jones.

A number people have read, been subjected to, and/or commented on portions of this work. Portions of Chapter 1 were read to the U. of A. Philosophy Colloquium--too recently for the current version to reflect the discussion that ensued. Tilman Lichter read and discussed Chapter 3 with me at length. Rich Campbell was kind enough to offer several pages of commentary on an earlier draft of material that has since been greatly transformed into Chapter 8, and Béla Szabados read portions of what are now Chapters 7 and 8. I presented part of Chapter 2 at the 1992 Alberta Philosophers' Conference in Banff, where I received helpful criticisms from F. J. Pelletier and Janet Sisson (whom I also have to thank for steering me down the path of Wisdom). An early version of Chapter 7, §§I-III was read and discussed by the U. of A. Philosophy Graduate-Student Workshop, where I received useful criticisms from Stephen Jones and (again) Tilman Lichter, and I am still digesting some thoughts shared with me when I read a later version at the 1993 Canadian Philosophical Association meetings in Ottawa. I also benefited from presenting portions of Chapters 2 and 3 to the Philosophy Department at York University in December of 1992.

Finally, I owe a special debt to Kathryn Harvey, who read and discussed with me much of what follows, and without whom all this and a great deal else would be very different for me.

M. J. H.
Edmonton, Alberta
October 8, 1993

Table of Contents

Introduction: Philosophy, Neurosis and Therapy 1

Chapter 1: Realism, Truth and Reference 12

I. Correspondence and Realism 12

II. Putnam's Model-theoretic Argument 16

III. Objections 20

IV. Reformulating the Argument: Some Interpretive Speculation 26

V. Two Accounts of Reference 33

Chapter 2: Realism and Scepticism 40

I. Brains in a Vat 40

II. Real Possibility 43

III. Realism and Scepticism: Epistemic Neurosis 48

IV. Metaphysical Revisionism 51

V. Realism? Anti-realism? Historicism? 59

VI. Truth 65

Chapter 3: Incommensurability and Conceptual Schemes 70

I. Conceptual Schemes 71

II. Davidson's Argument 77

III. Truth, Interpretability and the Principle of Humanity 84

IV. Translation and Indeterminacy 91

V. Interests, Activities and Incommensurability 97

Chapter 4: Relativism, Truth and Rationality 102

I. Relativism as a Response to Realism 103

II. How Relativism Defeats Itself 107

III. The Relativist as Realist: Cultural Essentialism 113

continued ...

Chapter 5: Relativism, Realism and Scientism 119

I. The Ethical-Political Argument 119

II. Realism and Scientism: Necessary Connections? 127

III. Realism, Scientism and Analytic Philosophy: Contingent Connections 133

IV. Realism, Scientism and Marxism: More Contingent Connections 138

Chapter 6: Self-Knowledge: Descartes, Hume and Kant 150

I. Descartes: On Mixing the Subjective with the Objective 150

II. Hume: The Illusion of Subjectivity 157

III. Kant: Self-Knowledge Is Not Self-Sufficient 161

Chapter 7: The Objective and the Subjective in First-Person Authority 169

I. Anti-Individualism: Putnam and Burge 169

II. Anti-Individualism and First-Person Authority 175

III. Real Possibility: A Problem for Burge 180

IV. An Alternative Picture 187

V. Asymmetry: Shared Experience and Trust 191

Chapter 8: Self-Interpretation and Self-Unity 200

I. Unarticulated Beliefs: Human Beings and Other Animals 201

II. Self-Knowledge and Self-Interpretation 208

III. Self-Ignorance and Self-Division 212

IV. Multiple-Selves 218

V. Self-Descriptions, Agency and False Unities 225

VI. Unity: Its Possibility and Its Desirability 231

VII. Concluding Therapeutic Remarks 237

Bibliography 240

Abbreviations

The following abbreviations are used for parenthetical references to frequently cited works. References to all other works are to be found in the notes.

A/B Kant, Immanuel. 1965. *Critique of Pure Reason.* Trans. Norman Kemp Smith. New York: St. Martin's Press.

_____. 1923. *Kritik der reinen Vernunft.* 6th ed. rev. Ed. Benno Erdmann. Berlin: Walter de Gruyter & Co.

CEOP Burge, Tyler. 1986. "Cartesian Error and the Objectivity of Perception." In *Subject, Thought and Context.* Ed. Philip Pettit and John McDowell. Oxford: Clarendon Press, 1986, pp.117-136.

CSM I/II Descartes, René. 1984. *The Philosophical Writings of Descartes,* 2 vols. Trans. John Cottingham, Robert Stoothoff and Dugald Murdoch. Cambridge: Cambridge University Press.

FMMS MacKinnon, Catharine A. 1982. "Feminism, Marxism, Method, and the State: An Agenda for Theory." *Signs* 7: 515-544.

FPA Davidson, Donald. 1984b. "First Person Authority." *Dialectica* 38: 101-111.

FR Feyerabend, Paul. 1987. *Farewell to Reason.* London: Verso.

ISK Burge, Tyler. 1988. "Individualism and Self-knowledge." *The Journal of Philosophy* 85: 649-663.

ITI Davidson, Donald. 1984a. *Inquiries into Truth and Interpretation.* Oxford: Clarendon Press.

LP Althusser, Louis. 1971. *Lenin and Philosophy and Other Essays.* Trans. Ben Brewster. New York: Monthly Review Press.

MEW Marx, Karl and Friedrich Engels. 1969. *Karl Marx und Friedrich Engels--Werke,* 42vols. Berlin: Dietz Verlag.

PI Wittgenstein, Ludwig. 1958. *Philosophical Investigations.* 2nd ed. Trans. G. E. M. Anscombe. Eds. G. E. M. Anscombe, R. Rhees and G. H. von Wright. Oxford: Basil Blackwell.

PMN Rorty, Richard. 1979. *Philosophy and the Mirror of Nature.* Princeton: Princeton University Press.

PP3 Putnam, Hilary. 1983. *Realism and Reason: Philosophical Papers, Volume 3.* Cambridge: Cambridge University Press.

R&T Devitt, Michael. 1984. *Realism and Truth.* Oxford: Basil Blackwell.

RRSK Barnes, Barry and David Bloor. 1982. "Relativism, Rationalism and the Sociology of Knowledge." In *Rationality and Relativism.* Eds. Martin Hollis and Steven Lukes. Cambridge, Ma.: MIT Press, 1982, pp.21-47.

RTH Putnam, Hilary. 1981. *Reason, Truth and History.* Cambridge: Cambridge University Press.

RVR Seller, Anne. 1988. "Realism versus Relativism: Towards a Politically Adequate Epistemology." In *Feminist Perspectives in Philosophy.* Eds. Morwenna Griffiths and Margaret Whitford. Bloomington and Indianapolis: Indiana University Press, 1988, pp.169-186.

SAAO Lugones, Maria C. 1990. "Structure/Antistructure and Agency under Oppression." *Journal of Philosophy* 87: 500-507.

STSS Hesse, Mary. 1980. "The Strong Thesis of the Sociology of Science." In *Revolutions and Reconstructions in the Philosophy of Science.* Bloomington and London: Indiana University Press, 1980, pp.29-60.

TA Baier, Annette. 1986. "Trust and Antitrust." *Ethics* 96: 231-260.

THN Hume, David. 1978. *A Treatise of Human Nature.* Ed. L. A. Selby-Bigge. Oxford: Clarendon Press.

WO Quine, W. V. 1960. *Word and Object.* Cambridge, Ma.: MIT Press.

WRPL Kripke, Saul. 1982. *Wittgenstein on Rules and Private Language.* Cambridge, Ma.: Harvard University Press.

Introduction: Philosophy, Neurosis and Therapy

My title takes its inspiration from John Wisdom's elegant essay "Philosophy and Psycho-Analysis."[1] The philosopher, Wisdom writes, is like the obsessional neurotic, who cannot leave her apartment without checking again and again to see whether she has turned off the lights, or whether the door has been properly locked. These doubts of the neurotic seem peculiar in much the way that the so-called "doubts" of the philosopher seem peculiar to the non-philosopher. I speak here of sceptical doubts: Do we have knowledge of the "external" world? Is mine the only mind in the universe? Can I even know my own mind?

But the philosopher's "doubts" are of not quite the same sort as the neurotic's. The philosopher "entertains" her doubts, as if occasional dinner-guests, and while they seem real enough for the duration of the meal, she doesn't let them linger once the party is over. She keeps them from interfering in her non-philosophical life. By contrast, the neurotic, says Wisdom, is moved to act. As little as she believes that she has left the light on, she still feels the need go back and check. "The philosopher doesn't," says Wisdom. "His acts and feelings are even less in accordance with his words than are the acts and feelings of the neurotic."[2] The philosopher--unlike the neurotic--often "doubts" and "worries" in a way that does not directly touch her life.

Is this diagnosis of "the philosopher" correct? If so, what is the etiology of her psycho-philosophical disorder? And is there some course of therapy that might alleviate or resolve her turmoil?

I believe that Wisdom's criticism is justified. The "philosopher"--and I mean this term neither to include all philosophers nor to exempt myself, necessarily--the philosopher, I shall argue, suffers from forms of "epistemic neurosis." That is, she is tempted by philosophical views that must "entertain" types of sceptical "doubt," if they are to be expressed, but she cannot take those doubts seriously, since so honouring the sceptic would undermine all available justification for the very positions that the philosopher wants to advocate.

The source of the philosopher's temptation in such matters, I shall maintain, is the conception of objectivity to which she is committed. Epistemic neurosis is the natural malady of those who indulge in the vice of dichotomizing the objective and the subjective. And this habit is supported by the related vice of regarding meaning and truth as non-intentional, non-normative notions. Indeed, the latter compulsion is but a contingent effect of the former, a consequence of applying a strong objectivism to the conceptual tools of semantics.

The therapy appropriate to such a disorder consists largely in painting a picture of an alternative, uncoloured by the hue of objectivity that prompts the neurosis. But since the temptation to dichotomize the objective and the subjective is strong, bolstered by the intellectual culture of our day, the mere exposition of another option is not enough. The drive toward the dualism of subject and object must be brought to consciousness wherever an epistemic neurosis

---

[1] Wisdom (1957).

[2] *Ibid.*, p.174.

shows its influence, so that this philosophical compulsion can face the "slow cure"[3] of critique.

The thoughts that I present here owe a debt of influence to the later work of Ludwig Wittgenstein. (Many interpreters of Wittgenstein will also see them as diverging notably from Wittgenstein's thinking.) This is most true of the accounts of objectivity and meaning central to my arguments. But it also holds of the metaphors of neurosis and therapy, which give thematic unity to the issues that I confront. In his *Philosophical Investigations* Wittgenstein makes some subtle and difficult remarks about the proper nature and task of philosophy. Among them is the following passage, which I quote at length:

> It is not our aim to refine or complete the system of rules for the use of our words in unheard-of ways.
>
> For the clarity that we are aiming at is indeed *complete* clarity. But this simply means that the philosophical problems should *completely* disappear.
>
> The real discovery is the one that makes me capable of stopping doing philosophy when I want to.--The one that gives philosophy peace, so that it is no longer tormented by questions which bring *itself* in question.--Instead, we now demonstrate a method, by examples; and the series of examples can be broken off.--Problems are solved (difficulties eliminated), not a *single* problem.
>
> There is not *a* philosophical method, though there are indeed methods, like different therapies. (PI §133)

"The philosopher's treatment of a question," he says later, "is like the treatment of an illness" (PI §255), and elsewhere he stresses the importance of providing a "*slow* cure" for "a disease of thought."[4]

In many ways my efforts fall short of the philosophical therapy that Wittgenstein recommends. I lack his clarity and simplicity of style, and it is not easy for my argument to be "broken off," since it relies less on examples and more on engagement with other philosophical texts. Moreover, there is a sense in which I take myself to be pursuing a "single problem" that manifests itself in different contexts. But what needs to be emphasized about my approach, as it needs to be emphasized about Wittgenstein's, is that thinking of philosophy as a kind of therapy does not comprise what some have termed a therapeutic "farewell to philosophy."[5] Nothing in the idea of therapy need imply the desirability of abandoning philosophy. Philosophical problems, like the real neuroses to which I compare some of them, have never been, and need never be, in short supply, though they may change with historical and cultural circumstances. Exactly what I find important about the problems I take up here is the persistent attraction of the views that I criticize, an attraction that I feel myself. The "real discovery," the discovery "that gives philosophy peace," is the one that lets us approach philosophical questions without being led into the philosophically neurotic, self-

---

[3] Wittgenstein (1981), §382.

[4] *Ibid.*

[5] Habermas (1987), p.306.

defeating ways that spring, *e.g.*, from a metaphysical picture of objectivity or from trying to give "theories" of meaning and truth, as though they were the objects of some *science*. Only when thinking does not rely on "doubts" that undermine philosophy itself can we deal properly with the philosophical questions that remain, and only then can we "break off" our list of examples (so that philosophical dispute does not "go on too long"[6]) without worrying that an ultimate ground of warrant might have been found, if only we had had the time.

And when Wittgenstein tells us that therapeutic philosophy "leaves everything as it is" (PI §124), he is not proposing an uncritical quietism, as Habermas seems to suggest,[7] or worse, an "academic sado-masochism, self-humiliation, and self-denunciation of the intellectual whose labour does not issue in scientific, technical or like achievements,"[8] as Marcuse complained. He is simply denying that there is a hidden essence of language, awaiting discovery by the scientifically-minded philosopher, and that there is any need or possibility of replacing our actual words and expressions with a pure, logical language that could sustain meaning apart from our possible interests and practices. Such criticisms do not do justice to the critiques of scientism and essentialism in Wittgenstein's later writings.[9] Indeed, I think it would not be amiss to view that work as a contribution to critical theory in much the sense upheld by thinkers like Marcuse and Horkheimer,[10] though it hardly wears such credentials on its sleeves.

It is not my chief task to interpret Wittgenstein's philosophy, but to the extent that my own work draws on his I want to insist on like

---

[6] Wisdom (1957), p.178.

[7] Habermas (1987), p.306.

[8] Marcuse (1964), p.173.

[9] They certainly betray no awareness of Wittgenstein's expression of intent in the Foreword to his transitional *Philosophical Remarks*:

> This book is written for such men as are in sympathy with its spirit. This spirit is different from the one which informs the vast stream of European and American civilization in which all of us stand. *That* spirit expresses itself in an onwards movement, in building ever larger and more complicated structures; the other in striving after clarity and perspicuity in no matter what structure. The first tries to grasp the world by way of its periphery--in its variety; the second at its centre--in its essence. And so the first adds one construction to another, moving on and up, as it were, from one stage to the next, while the other remains where it is and what it tries to grasp is always the same. (Wittgenstein, 1975, p.7.)

[10] Less "reckless" interpreters (and critics) of Wittgenstein than I are wont to balk at such suggestions. But the idea is not my pure invention. For hints of similarity see, for example, Lovibond (1983), Rosen (1983), Staten (1984) and Janik (1985). In general, I believe that taking Wittgenstein's anti-essentialism seriously requires not essentializing Wittgenstein's work.

provisos for my interests in the therapeutic aspect of philosophy. And while the present work is not primarily a work of radical theory either, its undertones will be clear enough. Like Horkheimer's "critical theorist," I deny that "social reality and its products" are "extrinsic" to philosophical work and that such matters belong solely to the sphere of "political articles, membership in political parties or social service organizations, and participation in elections."[11] Indeed, I take it as a constraint of adequacy on philosophical work that it display the real possibility of harmony with the concerns of critical theory, especially some recent work in philosophical feminism.[12] As Alasdair MacIntyre says, "All philosophy, one way or another, is political philosophy."[13] So, part of what I hope to achieve, even as I focus on the work of mainstream philosophers influenced by the analytical tradition of the twentieth century, is a clarification of concepts and problems that have an important role to play in radical theory. This aspect of my position is most evident in Chapters 5 and 8, but it is an underlying motivation throughout.

My investigation of the objective-subjective dichotomy and the epistemic neuroses that it provokes will fall roughly into three parts in which knowledge of the "external" world, knowledge of other minds and other cultures, and knowledge of ourselves will, respectively, provide the foci of discussion. Each of these aspects of human knowledge is a site of sceptical "doubt" and, so, a likely locus at which to encounter some form of epistemic neurosis.

I begin in Chapters 1 and 2 with an examination of metaphysical realism and the normativity of truth and meaning. My explorations here take the form of a consideration and qualified endorsement of the work of Hilary Putnam since the publication of *Reason, Truth and History*. In Putnam's writings the questions of realism, on the one hand, and meaning and truth, on the other, are closely intertwined, sometimes to the point of conflation. As a result of this and other unclarities, Putnam's ideas have been the subject of considerable misinterpretation and of criticism based on such misinterpretation. In Chapter 1 I examine his "model-theoretic" argument against the correspondence-theory of truth, as well as criticisms of the argument that have been made by a number of philosophers. To understand what Putnam is up to, we ought to recognize that the argument is directed not so much against metaphysical realism as against non-normative conceptions of meaning and reference. As such, its lineage can be traced to

---

[11] Horkheimer (1982), p.209.

[12] Here I am encouraged by the recent work of a number of feminist thinkers. See, *e.g.*, Frye (1983), Heldke (1988), Code (1991) --to name but a few. A number of philosophers have tried to draw connections between feminism and "pragmatist" thinking, with which my position has certain strong affinities. See, *e.g.*, Hanen (1987), Nielsen (1987), Heldke (1988) and Rorty (1991b). Explicit remarks about the relevance of Wittgenstein for feminism have been hard to come by, but there is increasing interest in the topic, as the 1992 and 1993 programmes for the Canadian Philosophical Association Annual Congress suggest. For earlier recognitions of the connections see Lovibond (1983) and Scheman (1983).

[13] MacIntyre (1987), p.398.

Wittgenstein's discussion of "rule-following" in *Philosophical Investigations*. So construed, the argument is one that I believe to carry considerable plausibility. Its point, in part, is that attempts to treat such concepts as truth, reference and meaning as though they were non-normative, as though they were "radically non-epistemic,"[14] issue forth in a radical indeterminacy of meaning and reference. This spectre of indeterminacy, whether of reference in the case examined by Putnam or of meaning generally in the "rule-scepticism" that Saul Kripke attributes to Wittgenstein, however, is exorcised by its own self-defeating nature. If meaning were indeterminate, then the problem of indeterminacy could not even be stated. More importantly, the alleged problem can be avoided by recognizing the normative character of semantic notions. To understand the meaning or reference of a word or an expression amounts to knowing how to use that word or expression correctly among many other expressions. In that case, semantic notions have an epistemic component--we could not be utterly deluded or ignorant about the meanings of our own words, since the correctness or incorrectness of our use is shown by whether or not we manage to get by in the world of other language-users. It is our practical interaction with the world and others that determines the meaning and reference of our linguistic utterances. The problem of indeterminacy does not arise here, though misunderstandings are always possible.

From the normativity of meaning I return to the question of realism in Chapter 2. Putnam's real argument against metaphysical realism turns on a particular understanding of the nature of Cartesian scepticism and on a distinction between two different accounts of objectivity (which I introduce in Chapter 1). The metaphysical realist holds that for something to be objective is for its existence and nature to be utterly independent of our abilities to know and describe it. By contrast, the "internal realist" regards objectivity as implying only that the nature and existence of a thing be independent of what anyone actually believes of it. The former, metaphysical account of objectivity, but not the latter account, tends to manifest a vulnerability to Cartesian scepticism. For something to be objective in the metaphysical realist's sense is for it to be possible that we have always been completely mistaken about its nature and existence.

Just as the indeterminacy of reference and meaning proves self-defeating in Chapter 1, Cartesian scepticism and the metaphysical realist's account of objectivity show themselves here to be self-defeating. For if the sceptic's scenario actually obtained, then the reference and meaning of our words would differ from what we take them to be and in quite a systematic way. This is because the things with which we would "interact"--if, indeed, we could be said to interact with anything--under such circumstances would not be objects and organisms in the "external" world, but the mere appearances of such things in the "virtual reality" of our deluded experience. So, if we cannot be in complete error about the meanings of our words, then we cannot really be deluded about the world in the kind of way that the sceptic suggests. A common objection to this sort of argument--that it shows only that we could not *express* our total delusion, not that we could not be so deluded--is raised and rebutted.

It is here that the notion of "epistemic neurosis" comes most clearly into play. For in charging that we might really always have

---

[14] Putnam (1978), p.125.

been brains in a vat, even if we could not say so, the metaphysical realist simultaneously affirms and denies the worries of the sceptic, and the conception of objectivity on which they rest. Seeing this requires making clear the sense in which the sceptic holds our total delusion about the world to be possible. The possibility of our being in total error about the "external" world, I argue, is treated by the Cartesian sceptic as a *real* possibility, a possibility whose actuality would explain some given phenomenon of which we are assumed to have knowledge. The sceptic's doubt turns on raising sceptical scenarios as explanations of the nature and variety of our experience and claiming that these scenarios would explain that experience just as well as the hypothesis of our epistemic contact with the "external" world would.

Because the metaphysical realist shares the Cartesian sceptic's account of objectivity, she ought also to treat the sceptic's doubts as real possibilities. However, doing so tends to raise deep problems for the epistemic status of her own positive programmes. If it were true that we could not express, *e.g.*, our actually being brains in a vat, then it would be equally true that the metaphysical realist could not express her positive proposals concerning the extent and limits of human knowledge. Once she has imagined herself a detached brain in a vat of nutrient solution, floating in ignorance for all time with a cluster of electrodes tickling its neurons, she finds it hard to get back the world that she has "doubted" away. So, she cannot take the sceptic too seriously, but she must. As a result, she often backslides in her pledge to the real possibility of sceptical delusion and to the account of objectivity on which her views rest. Thus, she sometimes dismisses the sceptic's scenario as just a logical possibility, or complains that the sceptic distorts knowledge or justification. But in doing so she pulls the rug out from underneath her own feet.

Scepticism is not something incidental to metaphysical realism, for a vulnerability to the sceptic's worries is what makes metaphysical realism the position it is. The sceptic is the realist's alter ego, but the two do not comprise a simple unity. The metaphysical realist also thinks that the sceptic is mistaken, that we do--or can--know something about the nature and existence of the world. She can only "entertain" her sceptical "doubts," because she cannot *live* with them. With such doubts always about, philosophy can get no peace--it is "tormented by questions which bring *itself* in question," questions which may "leave no room for the rational activity of philosophy" (RTH 113). But her philosophical dinner-party will fail without her guests, and so, she must risk having to throw them out, when they insist on staying the night. It is this concurrent need to admit and deny a thesis, especially a sceptical one, that I call "epistemic neurosis."

In this chapter I also interpret Putnam's controversial treatment of truth as idealized rational acceptability. The central claim of his view, I argue, is simply that truth is not "objective" in the sense used by the metaphysical realist. Rather, as I argue in Chapter 1, it is a normative, epistemic notion, and while it cannot be reduced to notions such as warrant and justification, it *is* internally related to them. We cannot understand truth without understanding concepts like warrant, and in coming to understand the one concept we also come to understand the other. All this is compatible with treating truth as objective in the weaker sense endorsed by the internal realist. I embellish on Putnam's idea by explaining how a proposition could be "ideally justified." Ideal justification, I suggest, is a property of any proposition for which doubt would be self-defeating. So, the claim

that we have not always been brains in a vat is ideally justified. Empirical propositions may approach such ideal warrant, but will not attain it. This does not prevent them from being true, but without ideal justification there is always the real possibility of falsehood.

Having discovered a likely cause of epistemic neurosis in the application of a dubious metaphysical conception of objectivity to the world beyond our senses and to meaning and truth, I turn in Chapters 3, 4 and 5 to our knowledge of other minds and other cultures. Here I attend to varieties of relativism, with the aim of showing that the relativist actually shares the metaphysical realist's conception of objectivity. Where the relativist differs from the metaphysical realist is in denying that there is any such thing as objectivity. But, despite her best intentions, the relativist's view not only presupposes that objectivity could be only what the metaphysical realist takes it to be, but also makes positive use of that account of objectivity, even as it appears to emphasize subjectivity.

In Chapter 3 I distinguish conceptual relativism from epistemic relativism. The conceptual relativist holds that there is no one set of metaphysically correct or objective concepts, that different cultures possess different conceptual schemes and that those schemes can be incommensurable with one another. This is to say that it can be really impossible for members of one culture to understand members of another culture, if their conceptual schemes are different enough. Drawing on the work of Donald Davidson, I argue that any reason that we could have for regarding holders of another conceptual scheme to be in principle incomprehensible to us would also be a reason for holding that they were not really language-users at all. The conceptual relativist is tacitly a metaphysical realist, not about the external world, but about other cultures and conceptual schemes, because to maintain that it might be really impossible for us to understand another culture is just to apply the Cartesian sceptic's account of objectivity to cultures and conceptual schemes. Conceptual relativism thus presents itself as a cultural version of the problem of other minds. The relativist, like the metaphysical realist, suffers from an epistemic neurosis, needing both to affirm and deny a metaphysical account of objectivity and the sceptical worries that accompany it.

It is also part of my argument here, however, to suggest that some of Davidson's claims are too strong. In particular, his argument is advanced as ruling out the very idea of a conceptual scheme, and it rests on the assertion that we have no conception of truth apart from the notion of translation--*i.e.*, that the concepts of truth and translation are internally related. My suggestion is that we choose another starting-point. Truth and translation are not internally related in the way that Davidson suggests, but ideas such as warrant and justification, on one hand, and understanding and interpretation on the other, are. We have no viable story of linguistic understanding without normative ideas like justification; linguistic understanding just consists in the ability to use words and expressions in *correct* ways. Such a revised version of Davidson's argument, I further suggest, counts against Quine's thesis of the indeterminacy of translation. But while the internal connection between warrant and understanding allows me to join with Davidson in rejecting conceptual relativism, it does not warrant the conclusion that there is no such thing as a conceptual scheme. Rather, my contention is that no set of necessary and sufficient conditions distinguishes one conceptual scheme from another. Nevertheless, because language and concepts have

their roots in practical affairs, in our interactions with the world and with each other, different sets of concepts can arise, suited to different ends, and these groupings can suffer from a kind of practical incommensurability. Where cultures have different practices geared to achieving or preserving different ends, it is unsurprising that practical disagreements can arise, which may never be resolved. But these disagreements always remain at their roots practical differences, not real impossibilities of mutual understanding.

In Chapter 4 I take up the less plausible claims of epistemic relativism. The epistemic relativist holds that both truth and rationality are relative to cultures. While these are logically separable theses, they are typically found together, because the relativist regards truth as reducible to standards of rationality and justification, which are in turn held to be utterly local to cultural circumstances. Focusing on arguments for epistemic relativism that have been made by Barry Barnes and David Bloor and by Mary Hesse, I show that these arguments often start from premises similar to my own in criticizing metaphysical realism. However, the relativist operates with the metaphysical realist's account of objectivity in mind and so, regards metaphysical realism and relativism, the objective and the subjective, as forming strict dichotomies. The epistemic relativist cannot argue for her position without using non-relativistic accounts of truth and justification, but to do so is to abandon the position that she seeks to take. She is once again driven to the epistemically neurotic behaviour of affirming and denying a metaphysical account of objectivity. Yet, the relativist is often unswayed by this criticism. I argue that if her position has any initial plausibility, it is because it tacitly relies on an appeal to the conceptual relativism examined in Chapter 3. As well, the epistemic relativist correctly recognizes the self-defeating nature of metaphysical realism and takes it as comfort that her position is no worse off than what she regards to be the only competing view. But these are not the only explanations for why the relativist is unmoved by the charge of self-defeat.

In Chapter 5 I consider what I call the "ethical-political argument" for relativism. This argument--sometimes recognized by critics of relativism, but seldom examined in any detail--starts with the charge that metaphysical realism is morally or politically corrupt, because it provides an ideological justification for imposing a single world-view on other, less powerful cultures and on disadvantaged groups within our own culture. Relativism, on this view, is the only alternative to metaphysical realism and ought to be embraced, if other cultures and disadvantaged groups are to receive the respect that they deserve. Focusing on versions of this argument due to Anne Seller and Paul Feyerabend, I show that, even if these charges against metaphysical realism were correct, they would constitute no argument for relativism. I then take up the largely unexplored issue of why the radical theorist would ever be tempted by such a line of reasoning. The explanation proves to be complicated, lying not only in the relativist's tacit acceptance of a metaphysical picture of objectivity, but also in her conflation of any kind of realism with *scientism*, the tenet that all and only science--modelled on a simplistic view of the physical sciences--counts as knowledge. Suggesting that my own position can meet the worries of the ethical-political argument, I make two tentative proposals regarding the persistence of the conflation of realism and scientism. First, metaphysical realism and scientism often keep company in contemporary

analytical philosophy, an indicator, maybe, of its Cartesian heritage. Secondly, realism and scientism are often closely intertwined in the tradition of Marxist theory, which has been an important influence on the discourse of recent radical theory. In both cases, however, the links between realism and scientism must be regarded as historically contingent, because realism--be it metaphysical or internal--does not entail scientism, and scientism does not entail realism. One might be an epistemic relativist and still endorse scientism.

Shifting my focus from knowledge of the world and of others to knowledge of the self, I argue in Chapters 6, 7 and 8 that psycho-philosophical disorder displays itself here once again. And because if *anything* is subjective, then seemingly the self is, we find here some particularly odd combinations of objectivity and subjectivity, even though the two terms retain their dichotomous construal all the while.

Chapter 6 is a brief historical survey, in which I examine mixtures of the objective and the subjective in the thought of Descartes and Hume and then discuss Kant's critique of those combinations. Kant, I argue, began to recognize what both Descartes and Hume missed--*viz.*, that the self need not be metaphysically objective, a Cartesian substance, in order to be said to exist, and that self-knowledge is neither infallible nor utterly irrelevant to the constitution of its object. A brief review of Kant's doctrine of inner sense and of his Refutation of Idealism bears out these claims. However, Kant could not quite free himself from dichotomizing objectivity and subjectivity, as is suggested by his continued talk of the self in terms of mental "representations" and his commitment to the idea that there must be a single set of pure concepts of the understanding--only one conceptual scheme with objective validity. Nor did he appreciate the import for these issues of the fact that human beings are language-users.

In Chapter 7 I take the linguistic turn that Kant did not and trace the influence of the objective-subjective dichotomy in Tyler Burge's recent writings on self-knowledge. Burge makes a subtle attempt to reconcile anti-individualism about intentional phenomena with a version of first-person authority. The former is the view that an individual's beliefs, desires and other intentional characteristics depend for their content on facts about her social and natural environment. Some version of this view is entailed by my claim of Chapter 1 that meaning is determined by the ways in which we interact with the things and organisms in our environment (together with the assumption that intentional phenomena can be correctly characterized in language). First-person authority is the view that an individual knows her own mind better than those of others and better than others know hers--that there is an asymmetry in the justification of knowledge-claims about the self and others. A strong version of first-person authority would be Descartes' contention that I can know the present occurrence and contents of my ideas infallibly, provided I am attentive enough. Burge purports to reconcile a weaker version of first-person authority with anti-individualism. I argue, however, that his claims rest on an improper understanding of Cartesian scepticism. Reiterating my account of Chapter 2, I show that Burge's version of first-person authority is itself very Cartesian and that he must adopt an even weaker version, while also abandoning metaphysical realism, if he is to be successful in reconciling anti-individualism with first-person authority. Once again the dichotomy of object and subject and the epistemic neuroses that it provokes display their influence.

For an alternative account of first-person authority, which avoids the pitfalls of metaphysical realism and related doctrines, I turn again to Donald Davidson, offering qualified support for his means of dealing with self-knowledge. Central to this view is a point that appears in Chapters 1 and 2--namely, that I cannot be utterly or mostly mistaken about the meanings of my own words. Self-knowledge, I suggest, is simply the capacity to articulate one's intentional phenomena, to *say* what one believes and desires, and this capacity goes hand in hand with general linguistic competence. However, whereas I cannot be largely in error about the meanings of my own words, I can often make mistakes about what others mean by their utterances. This, according to Davidson, is sufficient to explain the asymmetry of justification between knowledge of self and knowledge of others and, therefore, to account for first-person authority. But while I think that it is correct to recognize this asymmetry, I also argue that the asymmetry is a matter of degree and that there are two general ways in which it can be eroded and even eliminated completely in some cases. First, the asymmetry between self- and other-knowledge can be eroded in ways that Davidson--in his overemphasis of the *interpretation* of others--does not seem to acknowledge. Very often I do not need to resort to interpretation to understand the words and non-linguistic behaviour of another person. This is especially true, if we share a considerable amount of background-knowledge and experience. As well, I argue that, *ceteris paribus*, the asymmetry decreases in cases in which persons stand in relations of justifiable trust with one another. Using Annette Baier's account of justifiable trust, I suggest that a justifiable trust-relation tends to require neither ignorance of the intentional characteristics of the trusted, nor self-ignorance on the part of the truster, and that justifiable trust is a precondition of linguistic competence and, hence, of knowledge of self and others. The apparent circularity of this procedure dissolves, once we recognize that not all intentional attitudes require linguistic capabilities or self-knowledge. As I go on to argue in Chapter 8, non-human animals and human infants are capable of possessing beliefs and the like, provided they are capable of displaying normative behaviour.

Secondly, as Davidson partly acknowledges, psychological disorder and ideological delusion can cloud my knowledge of myself, so that what I need most for self-understanding is self-*interpretation*, sometimes with the help of others. I can face similar, but non-pathological, difficulties, should I be put in the position of having to make a difficult moral or prudential decision--knowing what to do requires knowing what I want to do and what I think I ought to do, what kind of person I think am. Drawing on my earlier discussion of Kant and on the work of Charles Taylor, I propose a way to supplement Davidson's account, in order to deal with such cases.

I conclude Chapter 8 with an investigation of the importance of my account of self-knowledge for an understanding of the nature of the self. Turning once more to the discourse of radical theory, I argue that a failure to come to grips with the dichotomy of the objective and the subjective has led to an exaggerated view of the fragmentation of the self. Focusing on writings by Louis Althusser and Maria Lugones, and drawing comparisons with the phenomenon of Multiple Personality Disorder, I show how it can make sense to speak of self-unity and in what way such unity might be desirable. While there is no set of conditions that is necessary and sufficient to distinguish one person or self from another, and while the Cartesian conception of the

self as substance with infallible access to its own intentional attitudes is mistaken, the possibility of a kind of unified self can be retrieved in the form of a "grammatical" unity of self-description. This is to paraphrase the claim of Chapter 1 that meaning is a normative, intentional notion as the claim that meaning is possible, only if there are creatures ("subjects") for whom things can have meaning--only, that is, if there are selves possessed of a certain degree of coherence and unity. Asking whether a unity of the self is desirable reduces to asking whether particular kinds of self-descriptions are desirable for the person to whom they are applied.

I do not regard the positions that I criticize here as simple errors of careless thinkers. As I have commented, I think that there is a real attraction toward the opposed views that I examine, an attraction extending beyond the fuzzy boundaries of philosophy to other academic disciplines and beyond the fuzzy boundaries of the academy, as well. Philosophy, in this instance, as in others, reflects its cultural circumstances at the same time as it contributes to them. I also believe that, as a result of this cultural embeddedness, the debates explored below are ones that, in Putnam's words, "we seem doomed to repeat ... (like a neurotic symptom), unless, perhaps, we can step back and offer a better (and deeper) diagnosis of the situation ..." (PP3 288). To be sure, there is no proxy for practical action in dealing with the contextual features that lead to neuroses--philosophical or other--and a better diagnosis is not necessarily a cure. But part of the distinctive character of therapy is that a new awareness of the etiology of one's disorder can, under the appropriate description, contribute to the beginnings of a recovery.

Chapter 1: Realism, Truth and Reference

I shall begin with the established philosophical project of trying to describe the relation between "language and the world." I take my cue here from the writings of a former metaphysical realist who has left the fold, Hilary Putnam. Putnam's views have been criticized by a number of commentators who have found his reasons for abandoning metaphysical realism wanting. If treated carefully his arguments can be rescued both from his own sometimes misleading pronouncements and from the criticisms of his opponents. His position is best thought of in relation to the later work of Wittgenstein, and it provides a basis for rejecting both metaphysical realism and the correspondence-theory of truth. I shall deal in this chapter with Putnam's critique of truth as correspondence and non-intentional theories of reference, turning to metaphysical realism proper in Chapter 2.

I. Correspondence and Realism

In *Reason, Truth and History* Putnam argues against a view that he describes variously as the "*externalist* perspective," a position whose "favourite point of view is a God's Eye point of view" (RTH 49), and "metaphysical realis[m]" (RTH 73). Metaphysical realism maintains that

> ... the world consists of some fixed totality of mind-independent objects. There is exactly one true and complete description of 'the way the world is'. Truth involves some sort of correspondence relation between words or thought-signs and external things and sets of things. (RTH 49)

In opposition to this view Putnam presents "the *internalist* perspective" (RTH 49).[1] Internal realism typically holds that

> ... there is more than one 'true' theory or description of the world. 'Truth', in an internalist view, is some sort of (idealized) rational acceptability--some sort of ideal coherence of our beliefs with each other and with our experiences *as those experiences are themselves represented in our belief system*--and not correspondence with mind-independent or discourse-independent 'states of affairs'. (RTH 50)

Although I shall make occasional mention of Putnam's insistence that metaphysical realism be characterized as holding that there is *one true theory*, I want to focus on the two other characteristics that Putnam includes. These are the notion of a "mind-independent" or "discourse-independent" world and the correspondence-theory of truth.

Putnam's inclusion of the correspondence-theory of truth under the heading of "metaphysical realism" has been challenged by some critics.[2] As he has responded, the point is not that metaphysical realism implies correspondence, but that the metaphysical realist

---

[1] --Also known as a "non-realist view" (RTH 57) or an "'internal' realist view" (RTH 60), and more recently as "pragmatic realism" or "realism with a small 'r'" (Putnam, 1987, p.17).

[2] See, *e.g.*, R&T 4.

usually finds correspondence appealing.[3] Nonetheless, to begin, I think that it is worth heeding Michael Devitt's maxim that we "[d]istinguish the metaphysical (ontological) issue of realism from any semantic issue" (R&T 3). How well these issues can be separated depends upon whether we accept or reject realism, but let me set that question aside for now and consider some theses that might reasonably be thought germane to the question of realism. I group them under three headings: metaphysical, semantic and epistemic.

Let's begin with metaphysics. The metaphysical realist typically holds that the nature and existence of the world are independent of our abilities to know or describe them, but (assuming that we do have empirical knowledge) that the world nonetheless exists and is quite determinate in its nature and, moreover, that its nature is to be thought of as identical with, roughly, the ontology of "common sense" or of the natural sciences. Embodied in this view are a number of distinct points that can usefully be distinguished.

First, consider the claim (1) that there is a reality whose existence is independent of our abilities to describe or know that reality. Kant thought that any philosophical position that rejected this thesis risked falling into incoherence. We need, he thought, something to serve as the *ground* of the appearances that he took to be empirical objects.[4] We can "*think* them as things in themselves" (B xxvi), and must be able to do so for "otherwise we should be landed in the absurd conclusion that there can be appearance without anything that appears" (B xxvi-xxvii). However, Kant's version of this thesis seems to leave open the possibility that the reality in question might be utterly indeterminate[5]--indeed, on one reading of Kant's first *Critique* this is crucial to the position that he embraces under the title "transcendental idealism." For it is characteristic of that view that "we cannot *know* ... objects as things in themselves" (B xxvi), since knowledge requires not just reason, according to Kant, but sensibility also. Were it the case that the thing or things in themselves were determinate, then Kant's position would resemble a kind of scepticism; it would amount to saying that there exists a determinate reality of which we cannot have knowledge. By contrast, an indeterminate thing in itself would seem in its very constitution to escape the possibility of determinate description. It is not an appropriate object for knowledge, and so, plausibility is lent to Kant's claim that empirical knowledge is knowledge of appearances.

By the same token, however, thesis (1) is not likely to satisfy the metaphysical or "transcendental realist" (A 369), because something that in its indeterminacy is intrinsically unknowable hardly seems what realists have in mind when they entertain questions about our abilities to know the external world. The metaphysical realist holds further that (2) there is a reality that is independent of our abilities to know and describe it and which is also *determinate*. It *is* some *way* and not some other ways. Its independence carries with it a sceptical threat. Unhampered by temporal restrictions, and given all our current methods and procedures of inquiry, plus any that we might

---

[3] Putnam (1990), p.30f.

[4] Bxxvi-xxvii.

[5] Devitt calls this "weak, or fig-leaf, realism." See R&T 234.

develop, our descriptions might still fail to fit the world. Put another way--we might right here and now be completely mistaken in thinking that the world is as it appears to us to be.

Putnam's metaphysical realist accepts both theses (1) and (2). Moreover, as Devitt observes, the realist often accepts a further thesis (3) that identifies the determinate nature of the independent world with roughly the ontology of "common sense" (R&T 15f) and, maybe, of science. This is just to say that realists generally hold that the world is best described as containing tables, chairs, people, dogs, mouse-lemurs etc. or as containing (subject to revision in the light of new evidence and theories) quarks, gluons, neutrinos, etc.

I want to attend briefly to the nature of the "independence" involved in these metaphysical theses. Devitt suggests that what is relevant here is that the world exist "*independently of the mental*" (R&T 14), and Putnam speaks of "mind-independence" and "discourse-independence." These turns of phrase fail to distinguish independence from our epistemic capacities, on one hand, from independence from our actual beliefs or knowledge, on the other hand.[6] This will be important, when we turn to Putnam's argument against metaphysical realism in the next chapter. The reason for Devitt's specification of "the mental" in his version of the independence-thesis is to distinguish realism from some kinds of idealism and some kinds of phenomenalism. "To say that an object has objective existence ... is to say that its existence and nature [are] in no way dependent on our epistemic capacities" (R&T 13). However, Devitt continues, "the unsensed sense data of some empiricists, and Kant's pre-conceptualized intuitions ... have objective existence in the above sense ..." (R&T 13). Thus, independence entails objectivity of a sort, but also more. I shall not always be as fastidious as Devitt in distinguishing independence from objectivity, but I believe that no part of my argument is thereby threatened. When I speak of independence and objectivity in this chapter, I generally have in mind the independence from our epistemic capacities that is at work in thesis (2).

Consider now two semantic claims of relevance to my discussion. I shall call these fourth and fifth theses (4) *semantic realism* and (5) the *correspondence-theory of truth*. Semantic realism is the claim that the truth-values of our beliefs and sentences are independent of our abilities to ascertain what those truth-values are. This is to apply a particular conception of objectivity to the concept of truth.

The correspondence-theory of truth is the further claim that what makes our beliefs and sentences true or false, in a way that is independent of our abilities to ascertain their truth-values, is that they determinately "correspond" with a determinate reality independent of our epistemic capacities. We can think of this correspondence as a kind of fitting,[7] in which a true sentence "fits" the world, much as a properly sized glove fits a hand or a key of the proper shape fits a lock. Devitt suggests that there are three crucial elements to correspondence: the structure of the sentence being considered, the referential relations of its parts to reality, and the "objective

---

[6] Putnam is more careful in *Representation and Reality*, when he distinguishes "Objective with a capital 'O'" from 'objective' "with a small 'o'." See Putnam (1988), p.109.

[7] Alvin Goldman uses this metaphor in Goldman (1986), pp.152-154.

nature of that reality" (R&T 28). The version of correspondence I shall consider adds the further claim that the reality in question is independent of our epistemic capacities. The correspondence-theory, then, implies theses (4), (1) and (2) and is usually connected with thesis (3). However, none of these theses, individually or collectively, implies the correspondence-theory.

All we need further to note are two epistemological theses, which are (6) that knowledge of the world is knowledge of a determinate, reality independent of our epistemic capacities--and, generally, knowledge of the ontology of (3)--and (7) that such knowledge is to be had in virtue of our possessing beliefs that are correspondence-true.

Now, Devitt maintains that what is central to realism lies with the first three theses, especially with the third.[8] What realism is primarily concerned with is the objective and independent existence of most "common-sense" objects and, maybe, the entities recognized by the natural sciences. Never mind, especially, the correspondence-theory of truth. On Devitt's view Putnam is guilty of a conflation made most explicitly in the writings of Michael Dummett.

We have Dummett to thank for the "colourless term 'anti-realism'",[9] though we should not blame him for the reckless abandon with which that term has come to be used. On Dummett's account, anti-realism opposes a view called (surprisingly enough) "realism," and realism is, in turn, a view that one can hold of a particular set of statements, *e.g.*, statements about theoretical entities, statements about the past or the future, or statements about material objects. A realist about a class of statements is someone who maintains of those statements that they "possess an objective truth-value, independently of our means of knowing it: they are true or false in virtue of a reality existing independently of us."[10] The anti-realist, by contrast, maintains of the "disputed" class of statements that their possessing truth-values depends upon our "means of knowing" them. Dummett's anti-realist

> ... insists ... that the meanings of these statements are tied directly to what we count as evidence for them, in such a way that a statement of the disputed class, if true at all, can be true only in virtue of something of which we could know and which we should count as evidence for its truth.[11]

These descriptions of realism and anti-realism run together just the sorts of issues that Devitt thinks ought to be kept quite distinct. Indeed, they violate another maxim that Devitt insists we ought to observe in discussing the issue of realism: "Settle the realism issue before any epistemic or semantic issue" (R&T 4).

But this maxim is no more than a persuasive definition of the terms of debate. Only if we presuppose the world to be independent of our epistemic capacities can epistemic questions be seen as secondary to the question of realism. Moreover, to the extent that propositional

---

[8] See R&T 11. *Cf.* Lewis (1984), p.231; BonJour (1985), pp.159ff.

[9] Dummett (1979), p.145.

[10] *Ibid.*, p.146.

[11] *Ibid.*

knowledge of the world must be expressed in language, issues of semantics cannot be thought of as irrelevant to the question of metaphysical realism, either, unless metaphysical realism is again presupposed. These points will be of importance later.

II. Putnam's Model-theoretic Argument

With the foregoing distinctions in mind I want to turn to Putnam's critique of the correspondence-theory of truth. To start it will be useful to sketch broadly the logic of his arguments for "internal realism." Chapters 1 and 2 of *Reason, Truth and History* offer two major arguments, one against the correspondence-theory of truth and one against the kind of independence and objectivity that Devitt classifies as realism. In Chapter 1 Putnam argues that if we assume what I shall call the "interactive conception" of reference to be correct, then a strong form of scepticism about the external world is self-defeating. Moreover, if that form of scepticism defeats itself, then the world lacks the kind of independence and objectivity advocated by the realist. Thus, since the interactive conception of reference is correct, realism is mistaken. In Chapter 2, Putnam argues against alternative conceptions of reference that derive from the correspondence-theory of truth. Because all such accounts of reference are ultimately disguised versions of a "magical theory of reference" (RTH 15), the correspondence-theory of truth fails. And since the only plausible alternative accounts of reference prove to be mysterious, we have further reason to embrace the interactive picture, which thwarts the independence and objectivity characteristic of realism.

It is important to bear the foregoing argumentational schema in mind, as brief as it may be, since many commentators seem to interpret the point of Putnam's arguments differently and are then led to make criticisms that do not seem interesting on my reading. He is accused, for example, of being a "sceptic about determinate reference,"[12] of holding that there is "something intrinsically indeterminate and open to reinterpretation in our linguistic *practice*,"[13] and of giving arguments that have "nothing to do with metaphysical realism."[14]

Careful inspection shows that Putnam's argument against independence and objectivity (theses 1-3)--his critique of the claim that we might all be brains in a vat--is compelling, if one takes the interactive conception of reference seriously. Indeed, even if we accept, with some realists, the claim that reference is to be *explained* by a "causal theory," Putnam's argument remains compelling, and claims to the contrary depend on *implicitly* taking a "God's Eye point of view." (I shall argue these points in Chapter 2.) Examination also shows that, although there are problems with Putnam's argument against the alternative accounts of reference--his so-called "model-theoretic" argument--it is a valid argument, and the problems from which it suffers are not those cited by some critics. In fact, it embodies an important insight that can be rescued from the choppy waters of model-

---

[12] Lewis (1984), p.225.

[13] Hacking (1983), p.105.

[14] Blackburn (1984b), p.568.

theory by turning to Wittgenstein's *Philosophical Investigations*. I begin with an early version of Putnam's model-theoretic argument.

In *Meaning and the Moral Sciences* Putnam presents his case as follows. Suppose that there is a determinate reality independent of our epistemic capacities--*i.e.*, suppose that metaphysical realism is correct. It is reasonable to assume that this determinate reality, which Putnam sometimes calls "THE WORLD,"[15] contains or can be broken down into infinitely many objects.[16] When we speak whatever language we speak, we take ourselves to be referring to these objects, or sets of them. That is to say that there is an intended interpretation of the words of our language, according to which certain parts of the world, rather than others, are picked out by certain bits of language.

According to the correspondence-theory of truth, it is in virtue of this determinate correspondence that what we say stands a chance of being true. Of course, it is not a %100 chance, because many of the things we have wanted to say about THE WORLD have proved--as far as we can tell--to be false, or at least unjustified, and this gives us some reason to suppose that we are now discussing THE WORLD in terms of a theory that is only partially or approximately true. Worse yet, metaphysical realism is vulnerable to a list of traditional sceptical worries. No matter how ideal our theory of THE WORLD may be--in terms of standardly recognized theoretical constraints, such as consistency, simplicity or predictive power, and in terms of particular operational constraints "within" the theory[17]--there is always a sceptic's chance that it is false. But just as the truth of any sentence is due to the satisfaction of the proper correspondence, so the falsehood of any sentence is due to a failure of that correspondence. Whether the key is the right key or not depends on whether or not it fits the lock.

Now, let us assume that we have such an ideal theory, T1:[18]

> Lifting restrictions on our actual all-too-finite powers, we can imagine T1 to have every property *except objective truth* --which is left open--that we like. E.g. T1 can be imagined complete, consistent, to predict correctly all observation sentences (as far as we can tell), to meet whatever 'operational constraints' there are ... to be 'beautiful', 'simple', 'plausible', etc. The supposition under consideration is that T1 might be all that *and still be* (in reality) *false*.[19]

We have here, then, a detailed version of thesis (3) above.

---

[15] Putnam (1978), p.123.

[16] I am uncomfortable with the claim that there are infinitely many objects, as opposed to "indefinitely many," but I set this aside.

[17] --Such as, "Probably, if red litmus paper turns blue when immersed in a solution, then that solution is a base."

[18] Putnam adds that our theory should "[say] that there are infinitely many things" (Putnam, 1978, p.125), but if it is reasonable to assume that THE WORLD has infinitely many pieces, then we could make this another theoretical constraint, along with simplicity, etc.

[19] *Ibid.*, p.125.

In order to allow for the sceptic's chance that T1 might be wrong, however, the metaphysical realist must have a way of singling out what the intended interpretation of the terms of the theory is, because--and this is what the model-theoretic argument is meant to show--for a consistent set of sentences that can be interpreted to be about objects in some infinite set of objects[20] there will always be some model, consisting of an interpretation and an infinite domain, according to which the terms of T1 *correspond* to the objects of that model's domain. But then, insofar as there is such a correspondence, the theory T1 will be true on that model. There will always be *some lock* that the key fits and some key that fits a given lock. A key is the *right shape* only relative to some particular lock. There is no key that is *absolutely* the right shape--apart from any particular lock.

Unless the metaphysical realist can give a reason for choosing one specific correspondence between the terms of T1 and one specific domain of objects--one way of "carving up" THE WORLD--then there is no reason to think that truth is independent of theory-interpretation. Moreover, as we shall see, Putnam holds that the realist *can* give no principled reason for choosing one interpretation over another.

This version of the argument emphasizes showing that an "ideal" theory could not be false, unless there is some principled method for restricting the many ways of interpreting the terms of a language. In *Reason, Truth and History* Putnam takes a slightly different tack, but with similar results. Suppose again that we have an ideal theory, as above, but let it also be the case that that theory is *true*, and not just true on any old interpretation. In other words, on the "intended" interpretation of the terms of T1, there is a correspondence between every sentence of T1 and some set of pieces of THE WORLD. But if that is the case, then we can tell for any sentence of the language what its truth-value is, simply because our theory correctly predicts the truth-values of all observation-sentences and tells us what theory-sentences are made true by what observations, etc. What Putnam argues is that even if a theory is true on the intended interpretation, there is still nothing privileged about that interpretation, because there will always be other interpretations that assign exactly the truth-values to all sentences that the intended interpretation does. Thus, correspondence on the intended interpretation does not *explain* truth, for there are other correspondences on other interpretations that could do the job just as well. In fact, Putnam claims that there are interpretations according to which the truth-value of every sentence will remain unchanged *in every possible world*[21] from the truth-value it is assigned under the intended interpretation.

The argument can be made more convincing by applying it to a few simple terms of English. Putnam has us consider the sentence 'A cat is on an mat'. We can interpret this sentence such that 'cat' may be taken to refer to cherries and 'mat' to trees in the actual world

---

[20] We can do the same thing for a finite world by taking our sentences to refer to objects in a finite domain. See *ibid.*, p.139n3.

[21] I treat possible worlds as, in some cases, a useful metaphor for illustrating intuitions about possibility and necessity. It should not be thought that such talk explains modality; rather, it always presupposes a prior notion of modality. See Hymers (1991). See also "Possibility and Necessity" in PP3 46-68 and Putnam (1992a), p.51.

without changing the truth value of the sentences in any possible world--provided we give the proper sort of interpretation to 'A cat is on a mat' at each world. We could, for instance, re-interpret 'A cat is on a mat' to mean 'A cat* is on a mat*', where we can define cat*hood and mat*hood in worlds where some cat is on some mat, some cherry is on some tree, no cat is on any mat, no cherry is on any tree, or where the non-contradictory conjuncts of the preceding conditions are true. Thus, we can give a simple visual display of Putnam's re-interpretation by constructing what we might call "transworld definition matrices" for cat*hood and mat*hood:

cat*hood

	Some cat is on some mat.	No cat is on any mat.
Some cherry is on some tree.	(a) x is a cat*, iff x is a cherry.	
No cherry is on any tree.	(b) x is a cat*, iff x is a cat.	(c) x is a cat*, iff x is a cherry.

mat*hood

	Some cat is on some mat.	No cat is on any mat.
Some cherry is on some tree.	(a) y is a mat*, iff y is a tree.	
No cherry is on any tree.	(b) y is a mat*, iff y is a mat.	(c) y is a mat*, iff y is a quark.

The conditions expressed outside the matrices specify, when taken in row-column conjuncts, classes of worlds. The sentences inside the matrices indicate the truth conditions at these worlds of the sentences, 'x is a cat*' and 'y is a mat*'. Brief scrutiny should reveal that in worlds of type (a) 'A cat* is on a mat*' is true, because some cherry is on some tree, but additionally, 'A cat is on a mat' is true, because some cat is on some mat. In worlds of type (b) 'A cat* is on a mat*' is true, because some cat is on some mat, but this also means that 'A cat is on a mat' is true. In worlds of type (c) 'A cat* is on a mat*' is false, because cherries cannot sensibly be said to be on quarks, but likewise, 'A cat is on a mat' is false, because no cat is on any mat. Therefore, in all possible worlds--*i.e.*, necessarily--a cat is on a mat, if and only if a cat* is on a mat*. By extending this method and playing a similar game with *all* of the sentences of the language, we arrive at the result that the intended interpretation is not needed to preserve the truth-values of our sentences in all possible worlds. We are left with an indeterminacy of reference that is radical in its proportions, and which points to the failure of determinate correspondence with a determinate world to explain truth. But surely, as Devitt suggests, it is the goal of a *theory* of truth to *explain* truth.[22]

---

[22] See R&T 27f.

III. Objections

The most common objection to Putnam's position is also the least compelling, but Putnam does not help matters by phrasing his position somewhat carelessly at times. The objection is that Putnam *assumes* the correspondence-theory of truth to be false "in order to show it is false" (R&T 190). Putnam, say his critics, begs the question.

The metaphysical realist usually thinks a simple correspondence-relation can be singled out--or at least, that what legitimate correspondences there are are few enough, and similar enough with respect to the domains of their models that there is a significant likelihood that 'cat' refers to cats and not to possessors of *cat*hood*. Thus, David Lewis writes,

> The correct, 'intended' interpretations are the ones that strike the best balance. The terms of trade are vague; that will make for moderate indeterminacy of reference; but the sensible realist won't demand perfect determinacy.[23]

Now, one popular way of saying what makes the intended interpretation *correct*, or more probably correct, is to say that on the correct model there is a causal link between the objects of that model's domain and the terms of the theory. Indeed, that causal link *determines* that our language be interpreted as it is interpreted on the model we have just mentioned to be "correct." Since reference is to *explain* truth, on this view, reference itself must be explained, and "[t]he hope here," says Devitt, "is for causal theories of reference" (R&T 27).

Putnam is at times unclear about how appeals to the appropriate causal chain between referent and word fall short. Consider what he says in response to a suggestion from Hartry Field[24] that there might be a physicalistic relation, *R*, such that "(1) x refers to y if and only if *x bears R to y*" (RTH 45):

> If reference is only determined by operational and theoretical constraints ... then the reference of '*x* bears *R* to *y*' is itself indeterminate, and so knowing that (1) is true will not help. (RTH 45f)

Prima facie, this response is not very compelling. Holders of a causal theory of reference are trying to provide something *in addition to* the "operational and theoretical constraints," and they argue that it is this additional constraint that makes it possible to fix reference,

---

[23] Lewis (1984), p.228.

[24] Field (1972).

where the initial constraints failed.[25] A later remark does nothing to assuage this impression of question-begging:

> [I]f I say 'the word "horse" refers to objects which have a property which is connected with my production of the utterance "there is a horse in front of me" on certain occasions by *a causal chain of the appropriate type*', then I have the problem that, if I am able to specify what *is* the appropriate type of causal chain, I must *already* be able to refer to the kinds of things and properties that make up that kind of causal chain. But how did I get to be able to do this? (RTH 66)

It is tempting to reply for the metaphysical realist that I got to be able to do this *by means of a causal chain of the appropriate type*. If people can refer determinately at all, then they have been doing so longer than anyone has been discussing a causal theory of reference. To suppose that a causal chain of the appropriate type cannot fix the reference of the terms of a correspondence-theory that explains reference by appeal to such causal chains is, as Devitt complains, to *presuppose* that a causal theory of reference is false.

Lewis mirrors Devitt's criticism of Putnam. Putnam's critique of causal theories can be seen as a conviction that any effort to produce the metaphysical realist's desired additional constraint--call it C--will result only in an extension of the original theory, T1, and that T1 will simply annex this extension, so that the terms of C will meet with the same indeterminacy as those of T1. Lewis replies,

> ... C is *not* to be imposed just by accepting C-theory. That is a misunderstanding of what C is. The constraint is *not* that an intended interpretation must somehow make our account of C come true. The constraint is that an intended interpretation must conform to C itself.[26]

Which interpretations of our language and theory are most likely to be correct will be selected for by some objective, determinate relation between THE WORLD and the terms of our theory--however interpreted.

---

[25] Field's hopes are more modest in this respect. He writes,

> *The reason why accounts of truth and primitive reference are needed is not to tack our conceptual scheme onto reality from the outside; the reason, rather, is that without such accounts our conceptual scheme breaks down from the inside.* (Field, 1972, p.373).

Field wants primarily to show that semantic notions can be physicalistically *reduced*, so that we can be sure that they are "compatible with physicalism" (*Ibid.*). In another respect this is not such a modest aspiration. One can admit that any particular occasion of reference will be instantiated by a physical system without supposing that a Nagel-reduction, whereby the laws of reference (?!) can be derived from the laws of physics, could be carried out.

[26] Lewis (1984), p.225.

But how are we to make sense of this? Both Devitt and Lewis accuse Putnam of stacking the cards in his own favour. The metaphysical realist offers an alternative constraint to fix reference; Putnam asks how the realist knows that the terms of that constraint uniquely refer. Putnam, thinks Lewis, is just a sceptic about reference:

> The rules of disputation sometimes give the wrong side a winning strategy. In particular, they favour the sceptic. They favour the ordinary sceptic about empirical knowledge; they favour the logical sceptic, Carroll's tortoise or a present-day doubter of non-contradiction; and they favour the sceptic about determinate reference.[27]

However, Putnam is no sceptic about determinate reference. Rather, he rejects so-called "causal theories of reference," where these must be understood as attempts to provide some sort of physicalistic *explanation* of reference by singling out "causal chains of the appropriate type." (More generally, he rejects *non-intentional* or *non-normative* accounts of reference, as will become clear later.) Putnam is not denying that we must be in some sort of causal contact with things in order to be able to refer to many of them. Indeed, as we shall see, that thesis is of crucial importance to his argument that we could not be brains in a vat. He simply rejects efforts to single out some one causal relation which is the reference-relation. There are many relations between our words and the world, but what matters for *reference* is the role that a thing plays in our day-to-day interaction with it. Indeed, the fact that we are inclined to think of cats, but not cats*, as *things* is an indication of how our interaction with the world influences our ways of categorizing it, and hence, referring to it. Reference arises from our practices and interests and cannot be reduced to physical relations.

In this sense a truth does not fit the world in the way that a key fits a lock. The latter "fitting"-relation *is* a matter of external, physical relations, whereas the former is internal. If I bend my key or heat it sufficiently, it will not fit the lock any more. But it will still be the same key, and I will still be trying to open the same lock. If burglars tamper with the lock, then I will not be able to insert my key any longer. But it will still be the same lock, and I will still be trying to open it with the same key. By contrast, if I change my sentence--say, by adding or deleting a word--it will no longer be the same sentence, and it may or may not still "fit" the world. 'The cat is on the mat' can be changed by substituting four words, but "it" is no longer the same sentence when "it" has been changed into 'Some cherry is on some tree'. If I change the portion of the world that my original sentence described--say, by pulling the mat out from under the cat--then it will no longer be the same "fact" or "state of affairs" to which I originally referred. If I know what the words I use mean, then there is no question of unlocking some hidden relation between those words and their referents. If I know what my words mean, then I cannot be surprised as I can when I learn that my lock has been changed or that my key has been bent out of shape. The

---

[27] *Ibid.*, p.225.

fitting-relation between a sentence and the world is an internal relation--the true sentence *says* what aspect of the world it fits.[28]

The Devitt-Lewis charge of question-begging also rests on a misreading of Putnam's argument (a misreading which, it must again be acknowledged, Putnam encourages at times). The problem for causal theories of reference is not simply that 'causes' is itself a victim of the indeterminacy of reference--that is, indeed, true only if causal theories of reference are false. The problem is, rather, that for any relation between the terms of our language and the objects of some domain, there will always be countless other relations between the terms and the objects of other domains. So, although there may be some causal relation between 'cat' and cats, there will also be a causal* relation between 'cat' and cats*. And even if "'causally related' is 'glued to one definite relation' [*viz.*, a causal relation] by causal relations, not by metaphysical glue'" (R&T 189), it is equally the case that 'causally related' is glued to some other definite relation (a causal* relation) by causal* relations.[29] When Putnam says, "how 'causes' can uniquely refer is as much of a puzzle as how 'cat' can,"[30] he is being sloppy. But when he points out that there is more than one relation that constitutes a model of our language--maybe causal relations constitute one such model--he is pointing out that the metaphysical realist owes us an *explanation* of why it is this particular relation that constitutes the "correct" model rather than some other relation. To the extent that no such explanation can be given, the "causal theory of reference" proves simply to be a "magical theory of reference."[31]

Once we have specified for all the terms of our language the "intended interpretation," we have specified whatever set of causal relations it is that is supposed to apply generally to the terms of our language in such a way that they refer to the objects in the domain of the intended model. If we have not, then that set of causal relations does not constitute the reference-relation.[32] Adding that this relation around which the intended model is built is a *causal*

---

[28] I draw here on Hacker (1992), p.267. For related criticism of Goldman's use of the "fitting"-metaphor (Goldman, 1986, pp.152-154) see Nielsen (1988). Nielsen does not remark on the difference between external and internal relations of fitting, but his criticism presupposes that Goldman's fitting-relation is an external one.

[29] "[E]ven if it is the case that the word 'reference' is connected to the relation R_{17} by R_{17} itself, it will also be the case that the word 'reference' is connected to the relation R_1 by R_1 itself, to the relation R_2 by R_2 itself ..." (PP3 296).

[30] Putnam (1978), p.126.

[31] The reading of Putnam common to Devitt and Lewis is identified in LePore and Loewer (1988).

[32] I am assuming here that once we have listed exhaustively all the ordered pairs of relata we have described the relation. If it be desired that causal relations be defined across possible worlds, that should pose no problem in principle, given that we are already assuming THE WORLD to contain infinitely many pieces.

relation, is merely to give a name to the model. It tells us nothing about why *that* model is correct and others are incorrect. Richard Rorty summarizes Putnam's point succinctly:

> [N]o matter what nonintentional relation is substituted for "cause" in our account of how the things in the content reach up and determine the reference of the representations making up the scheme, our theory about what the world is made of will produce, trivially, a self-justifying theory about that relation. (PMN 295)

Actually, this is not quite enough--at least, it would not satisfy Hartry Field, who insists that a proper reduction of reference to a physicalistic level would have to be something more than a list. Using the example of chemical valence, he writes that an alleged reduction of valence, which merely eliminated the word 'valence' by pairing every known element and compound with an integer would be no genuine reduction at all.[33] What is needed is a reduction--or an "approximate reduction"[34]--of the "laws" of reference to succinctly worded causal laws. However, to just the extent that we can imagine this to be possible, we can likewise make sense of finding a reduction of the "laws" of reference to causal* laws, and nothing presents itself as a reason for choosing one of these purported reductions over the other. We are back where we began.[35]

Now, I have been treating Devitt and Lewis together, because each makes the same criticism of Putnam, but unlike Devitt, Lewis does not think that a causal chain of the appropriate type is sufficient to explain reference.[36] Lewis' ploy is a clever one, if occult. Rather than attempt to select one amongst many relations as *the* relation of reference, Lewis attempts to restrict the class of models of our language by limiting them according to their domains. That is to say, Lewis thinks that some ways of breaking THE WORLD into pieces are better than others and that only models that interpret the terms of our language and theory by assigning them to these particular pieces--or pieces very much like them--are eligible to be intended models.

His strategy for developing this constraint is outlined in "New Work for a Theory of Universals" where he asserts that only objects having

---

[33] See Field (1972), p.362ff.

[34] See *ibid.*, p.374.

[35] This is still not a sufficient response to Field, since he seeks only a reduction of reference, internal to a particular theory. So, he might reply that I still have not shown anything wrong with such an internal reduction. I deal with this in Section V below.

[36] Indeed, it would be a surprise if Lewis thought otherwise. His theory of modality is premised on the existence of infinitely many "possible worlds" from which we are all causally and spatio-temporally isolated. These worlds somehow manage to *represent* possibilities, including possibilities for actual individuals, despite this lack of causal contact. As well, we can have knowledge of these complete physical systems, a feat that is mysterious on a causal theory of reference. See Lewis (1987). See Hymers (1991) for a critique of his views.

relatively *natural* properties are things to which we can refer. These natural properties are intimately connected with universals:

> Natural properties would be the ones whose sharing makes for resemblance, and the ones relevant to causal powers. Most simply, we could call a property *perfectly* natural if its members are all and only those things that share some one universal. But also we could have other less-than-perfectly natural properties, made so by families of suitable related universals. Thus we might have an imperfectly natural property of being metallic, even if we had no such single universal as metallic, in virtue of a close-knit family of genuine universals one or another of which is instantiated by any metallic thing.[37]

So, it seems that THE WORLD is composed basically of perfectly natural properties, and out of these properties we can construct somewhat less natural properties. The more dissimilar are the constructing natural properties, the less natural will be the constructed properties. Anything that possesses wildly unnatural properties is immensely more difficult to refer to than something that possesses only relatively natural properties. Indeed, we may be unable to refer to it at all. Thus, any interpretation of our language that suggests that our words refer to a great many unnatural things is much less likely to be a nearly correct interpretation. On this view, cat*hood and mat*hood are extremely unnatural properties. Consequently, any model whose domain contains cats* and mats*, rather than cats and mats, is a far less eligible model of our language.

This proposed constraint has the virtue of satisfying realist intuitions about what sorts of things there are in THE WORLD, but Lewis does not explain *how* it is that natural properties succeed in fixing reference. Indeed, the very notion of a natural property begins to look rather mysterious on closer examination. Consider the sorts of objects in the domain of the intended model, M. There are relatively natural things like cats and metals and even more natural things like protons and sub-nuclear particles. But now, consider the domain of another model, M*. According to M*, there are many relatively natural* things like cats* and metals* and even more natural* things like protons* and sub-nuclear* particles. As our language is interpreted on M*, we cannot refer to unnatural* things, like cats, metals, protons and sub-nuclear particles. Natural* properties--by the way--"would be the ones whose sharing makes for [resemblance*] and the ones relevant to [causal*] powers." To suppose that natural properties *just are* the ones relevant to fixing reference is, again, to adopt a theory of reference that succeeds by magic. "What Lewis' story claims is that the class of cats cries out for a label, while the class of cats* does not cry out to be named."[38]

---

[37] Lewis (1983), p.347.

[38] "A Defence of Internal Realism" in Putnam (1990), p.38. Putnam anticipates my reply to Lewis in *Reason, Truth and History* in response to the objection that cat*hood is an *extrinsic* property (see RTH 37). His solution to the problem relies on the fact that extrinsic properties can be defined in terms of intrinsic properties, much as unnatural

IV. Reformulating the Argument: Some Interpretive Speculation

I indicated earlier that I saw a problem for Putnam's "model-theoretic" argument against the correspondence-theory of truth and causal theories of reference. It is a difficulty quite distinct from those I examined in the preceding section. Simply put, Putnam relies in his argument upon the "Skolem-Löwenheim Theorem," which is a theorem about sentences of first-order logic. However, it is unclear--indeed, unlikely--that English or any other natural language can be adequately formalized using only first-order logic.[39] Ian Hacking observes that this worry can arise in a number of ways:

> No one has ever shown that the commonplace language of physicists can ever be squeezed into a first-order format. ... [O]rdinary English primarily deploys second-order quantifiers. ... Much common speech involves what are called indexicals. ... Only dogma could insist that [such common speech] is expressible in first-order logic.[40]

This criticism seems to me quite devastating for the form in which Putnam has given his argument. However, I also think that his argument is better thought of as a Wittgensteinian argument. He raises the comparison himself, indicating that he "saw a connection" between the Skolem-Löwenheim Theorem and "some arguments in Wittgenstein's *Philosophical Investigations*" (RTH 7). Since this remark is not explained, and the comparison is obscured by Putnam's more frequent references to Kant, we are left to speculate on what connection Putnam saw and why he thought it relevant. I offer the following speculation.

In his *Philosophical Investigations* Wittgenstein devotes many passages to exploring the motivation and coherence of the suggestion that "if anyone utters a sentence and *means* or *understands* it he is operating a calculus according to definite rules" (PI §81). After a brief criticism of this view, focusing on misunderstanding an explanation and the purpose-relativity of "exactness," Wittgenstein

---

[38] properties are defined in terms of natural properties. Granted, the intrinsic-extrinsic distinction is not the same as the natural-unnatural distinction--that is why Lewis gets philosophical mileage from natural properties in places where intrinsic properties have failed--but what matters is that both intrinsic and natural properties are the ones we take to be primitive or basic. Putnam takes the so-called extrinsic properties as basic, arriving at definitions of cathood and mathood, according to which cat*hood and mat*hood are primitives, and "it is 'cat' and 'mat' that refer to 'extrinsic' properties, properties whose definitions mention objects other than x ..." (RTH 38). Similarly, if we take cat*hood as basic, it will, as-sessed from the M*-model, seem quite natural*, while cathood will seem unnatural*.

[39] Calvin Normore has suggested to me that we might view the argument as a *reductio* of the view that language can be captured by first-order logic. It is unclear to me that natural languages can be adequately formalized--period.

[40] Hacking (1983), p.105.

switches to discussions of the nature of philosophy and the form of a proposition, only to return later to linguistic rules.

His examination has been treated by some commentators as raising the spectre of "rule-scepticism"[41]--the worry that since a rule can be interpreted in indefinitely many ways, we are never justified in thinking that others use words in the same way that we do, or even that our own present uses are not guided by a rule different from that which guided our past uses. Thus Saul Kripke writes:

> The sceptic claims (or feigns to claim) that I am now misinterpreting my own previous usage. By 'plus', he says, I *always meant* quus; now, under the influence of some insane frenzy, or bout of LSD, I have come to misinterpret my own previous usage. (WRPL 9)

"Quus" is an alternative interpretation of "plus," according to which x quus y = x plus y, provided that x and y are each less than 57. Otherwise, x quus y = 5. The problem, as Kripke sees it, is that I tend to suppose that in learning the rule for addition I have somehow been given, or have given myself, instructions for dealing with any possible case of addition. Yet, I seem to understand this rule, these instructions, only on the basis of a finite, relatively small number of particular instances that could as easily count as instances of some other rule, *e.g.*, the quus-rule.[42] There might be, the sceptic infers, no fact concerning what rule I was following in the past. And I am no better off in the present, for the present instance is just one more in a finite number of cases and does not differ in this respect from past cases. The answer that I give now in computing '68+57' is compatible with my following indefinitely many different rules, other than the rule for addition. There might be, the sceptic insists, no fact of the matter about what I mean by 'addition', '68+57', 'doing the same', or anything else. "It seems," says Kripke, "that the entire idea of meaning vanishes into thin air" (WRPL 22).

Presenting this as a Humean sceptical paradox, Kripke argues that Wittgenstein can be understood to offer a Humean sceptical solution--one which concedes that "the sceptic's negative assertions are unanswerable" (WRPL 66), but which argues that "our ordinary practice or belief is justified because--contrary appearances notwithstanding--it need not require the justification the sceptic has shown to be untenable" (WRPL 66). On Kripke's reading, Wittgenstein concedes the force of rule-scepticism, but offers an alternative justification for our ordinary belief that we generally continue to follow the same rules, when we are engaged in solving problems of arithmetic, and when we make use of myriad linguistic expressions. Wittgenstein's Kripkean alternative is to say that it is possible to follow a rule only insofar as one is a member of a linguistic community--or rather, only insofar as one is *thought of* as a member of a community. "Our community can assert of any individual that he follows a rule if he passes the tests for rule following applied to any member of the community" (WRPL 110). Call this the "community-view."

---

[41] See, *e.g.*, Kripke (1982) and Fogelin (1976).

[42] See WRPL 10f.

We need further clarification of the community-view. What Wittgenstein is *not* claiming, says Kripke, is that whether or not a person is following a particular rule depends upon whether or not her community says she is:

> Wittgenstein's theory should not be confused with a theory that, for any *m* and *n*, the value of the function we mean by 'plus' *is* (by definition) the value that (nearly) all the linguistic community would give as the answer. (WRPL 111)

This view would entail that "125 is the value of the function meant for given arguments, if and only if '125' is the response nearly everyone would give, given these arguments" (WRPL 111).

Kripke's community-view is, rather, based on rejecting the idea that meaning can be identified with truth-conditions. An important part of Wittgenstein's account of meaning, says Kripke, is to see the meanings of assertions as given by their *assertibility-conditions*--conditions under which we are licensed in asserting particular propositions. Wittgenstein, on Kripke's reading, tries to answer the rule-sceptic by focusing on the assertibility-conditions for statements *about* meanings and rules, with the proviso that those conditions have some role to play in our lives. Thus, we are told:

> All that is needed to legitimize assertions that someone means something is that there be roughly specifiable circumstances under which they are legitimately assertible, and that the game of asserting them under such conditions has a role in our lives. (WRPL 77f)

These assertibility-conditions, thinks Kripke, are bound up with the fact that the members of a community tend to agree in what constitutes a correct application of a rule to a new case. We are justified in attributing to you the ability to carry out addition, if in sufficiently many cases you give the same answer as we would, or at least appear to be applying the same procedure as we would, even if you are making a mistake. We will not be justified in withdrawing our assessment, unless you consistently give an answer that we would not give, or unless your mistakes are so bizarre as to suggest that you are simply applying a different procedure, or no procedure at all.

But does this really deal with the sceptical worry that Kripke attributes to Wittgenstein? Shouldn't the rule-sceptic simply respond that we are justified in attributing to you the ability to do addition, only if we are not equally justified in attributing to you the ability to do quaddition, or some still more exotic "-ition?" Precisely the same conditions which, according to Kripke, would justify our assertion that you have followed the rule for addition in the past would justify another assertion that you have followed the rule for quaddition in the past. And as long as we contingently never face an instance like '68+57', we will find that you are inclined to give just the same answers as we are.

I suspect that it is this objection that Kripke hopes to anticipate and deflate by adding the specification that the particular "language-game" in question--attributing certain concepts to others--must have "a role in our lives" (WRPL 78). Attributions of an understanding of quaddition do not normally have any role to play in our lives, whereas attributions of an understanding of addition do. However, should

someone begin producing solutions radically different from the ones that we are inclined to give, then attributions of the concept "quaddition" would have a point and, hence, a justification.

But, the sceptic might persist, what of alternative rules whose different consequences might never display themselves, *e.g.*, a rule that was like the rule for addition, unless one or more of the numbers being added was greater than 10^{100}, in which case, as with quaddition, the solution of the sum would be 5? The response, of course, is just the same as before. We would not be justified in attributing the 10^{100}-rule (call it "googoldition"), because such an attribution would play no role in our lives. It would make no difference to us for practical purposes whether someone followed addition or googoldition, and we would be justified in attributing googoldition to someone, only if we had some reason to think that she would produce unexpected solutions under some set of circumstances. But it is unlikely that we would ever have such a reason.

Notice that this reply does not rule out the hypothesis of the rule-sceptic. All that it does is to say that we have no reason to believe that the sceptic is right. On this view the sceptic's proposal remains a possibility--and not just a logical possibility, but what I shall call in Chapter 2 a "real possibility." It is a possibility whose actuality would explain the actual solutions that a person gives to particular addition-problems, but its probability is regarded as being lower than that of another real possibility--that the person under study is doing addition. However, the philosophical problem[43] for Kripke's treatment lies in the very fact that it allows even this much possibility to the sceptic's proposal. And the source of this problem is Kripke's conflation of normative rules with empirical regularities. The rule-sceptic's possibility is raised as an *explanatory hypothesis* concerning linguistic behaviour, but linguistic or grammatical rules are *norms of linguistic conduct*, not explanatory hypotheses.

This criticism, if correct, would come as something of an irony, because the conflation of the normative and the empirical-explanatory is something against which Kripke himself warns in elucidating the sceptical paradox. Consider one tempting response to the rule-sceptic. The sceptic claims that, because I have not been given explicit instructions about how to apply a rule to the indefinitely many new cases that can arise, there are indefinitely many alternative rules that I might be following. The *dispositionalist* responds that this worry stems from focusing on occurrent mental states. Clearly, there is no moment at which I receive "in a flash" (PI §§138, 191, 197, 318, 319) an explicit understanding of a rule in the form of explicit instructions for every new case. However, what makes it the case that I follow this rule, rather than some other, is that I have a disposition to respond to new cases in *this* way, rather than *that*.

As tempting as this response may be, it is vulnerable to precisely the same objection that the sceptic raised before.[44] Any disposition that I have will display itself only in a finite number of cases, and these cases are compatible with my having any one of indefinitely many other dispositions. So, I am not warranted in thinking that I have any particular one. Further, the idea that my being *justified* in my answer

---

[43] I think it is a mistake to ascribe it to Wittgenstein, too.

[44] See WRPL 23.

to each new case by my being disposed to answer thus-and-so, rather than such-and-such, turns on a faulty conception of rules and justification. Being justified in the claim that I understand the rule for addition turns on more than my being disposed to answer in a certain way--though if I understand how to perform addition, then I might well be disposed to respond in that way. My understanding of a rule cannot be reduced to my disposition, because there are indefinitely many different dispositions that I could be said to have, each compatible with my past behaviour. And if the rule that I follow is given by my disposition, then we can make no sense of my making mistakes. Whatever I am disposed to say will be correct. Only if we have some other criterion for identifying the relevant rule, can we say whether or not my disposition conforms to it.

Yet how, we might ask, does Kripke's account of Wittgenstein's alleged sceptical reply avoid this same problem? On that view, our justification for saying that someone understands addition is that she has the same disposition as we have. But, as Gordon Baker and Peter Hacker ask, how do we know what disposition *we* have? "Given that no one previously ever added 57 and 68, how do we know that our present community-wide inclination to answer '125' accords with what we would have been inclined to say, had we previously been asked what 57+68 is?"[45] We know only what we have been inclined to answer in the past, not what we would be inclined to answer to future cases that we have not yet considered. Only if we have some further criterion for identifying the rule that we follow, can we say whether or not her response is the same as ours, because criteria of sameness are internal to particular rules. It is plausible that Wittgenstein has this in mind in two successive remarks from the *Investigations*:

> The word "agreement" and the word "rule" are *related* to one another, they are cousins. If I teach anyone the use of the one word, he learns the use of the other with it.
>
> The use of the word "rule" and the use of the word "same" are interwoven. (As are the use of "proposition" and the use of "true.") (PI §§224-225)

What counts as agreement from one case to the next is precisely what is determined by the rule--or, more carefully, by *our use* of an expression or a paradigm *as a rule*, as a guide for correctness, a standard of criticism, etc. Whether one case is the same as another is just the issue of whether or not the rule applies to it, and understanding the rule just is being able to recognize when and how to apply it. This is because rules have a normative function; they are not simply summaries of empirical regularities. They are not simply records of what has been done, but guides to what should be done, or to what counts as having been done and what does not.

There is something peculiar about the case that Kripke cites, a peculiarity that stems possibly from its being a mathematical case. Would the sceptic's worries appear even slightly credible if we applied them to, *e.g.*, baking? Should I worry that I might actually be following rules for "quaking," that although in the past I have always combined my pumpkin-pie ingredients, poured them in a pastry-shell and

---

[45] Baker and Hacker (1984a), p.37. I am much indebted to Baker and Hacker's discussion.

set the pie in the oven, I might suddenly decide that I should combine the ingredients, pour them in the shell and hurl the messy combination from my balcony at nonplussed passers-by?[46]

On Baker and Hacker's reading, Wittgenstein is not advocating a new kind of scepticism about rules, which could be given no more than a surrogate "sceptical" solution, but is exploring what it means to follow a rule, rejecting various tempting proposals along the way, because they lead to *absurdities*. And typically the absurdities are of the order of the consequences of rule-scepticism. Rather than embrace the sceptic's doubt, Wittgenstein takes the absurdities that follow from sceptical doubt to be reasons for rejecting the pictures of rule-following that produce such doubt.[47]

Why should we see the results of rule-scepticism as absurdities? If the sceptic is right, then there is nothing more to be said. If there is no way of deciding what rules we follow, then there is no way of deciding what our words mean. And if there is no way of deciding what our words mean, then there is no way of deciding what the sceptic's words mean. The sceptic's doubt is self-defeating, because if it is true, then it is meaningless; it must be false even to be expressed. Unless we actually follow specific rules, the sceptical question cannot arise, because we could not even *identify* relevant cases for assessment, unless we *already* followed some determinate rule. As I suggested, what counts as a case or as "sameness" from case to case is internal to the rule in question.

The whole threat of rule-scepticism acquires whatever force it may appear to have from the same conflation that Kripke's "sceptical solution" involves. This should not be surprising, since what makes a solution a "sceptical" one is the fact that it accepts all of the sceptic's premises. When we are confronted with empirical phenomena that we hope to regiment under statements of laws and regularities, we discover that there are many different hypotheses that would account for the same phenomena, given our background theories. But--setting aside the question of how we ought to deal with *that* philosophical issue--the situation is different when we are considering intentional, normative "phenomena." Rules are not just inductive generalizations. The idea that there are many rules that I *might* be following in using my words as I do is just the idea that I might be completely mistaken about the meanings of my words. And if that is the case, then I cannot understand the sceptic's doubt.

One might, of course, try to deny that we use language and insist that we merely seem to do so, that our words merely seem to be meaningful. But if I seem to use words, then I seem to use them in some particular ways and not in others. Otherwise, there is no content to the notion of seeming. Seeming to use words and to understand them

---

[46] I am inspired here by Cavell's bizarre "qualking"-example and his connection of sceptical doubt with the desire for "some assurance against a certain fear of going mad, or being defenceless against the charge of madness." However, the overall reading of Wittgenstein to which I am sympathetic differs from his. See Cavell (1990), p.85f.

[47] The point is also made by Malcolm, who does not accept Baker and Hacker's "individualistic" reading of Wittgenstein on rule-following. See Malcolm (1986), pp.154-181. For his critique of Baker and Hacker see Malcolm (1989). For reply see Baker and Hacker (1990).

is not like having a headache; it is not a brute sensation that happens to accompany some instances and not others. There must be cases that count as seeming to be, *e.g.*, cases of addition, and cases that do not, if I am to be said to seem to do addition. But if there are such cases, then meaning has been smuggled in through the back door. It can *seem* to me that I do addition in some case in which I do not *actually* do addition, only if I understand the concept of addition. It can only seem to me that I speak a language, if I understand the concept of a language. And to understand these concepts is to follow rules and to understand a language.

But then, what does understanding a rule consist in? How should we characterize this normative "phenomenon" beyond saying that it must be different from mere empirical regularity? The pictures that lead us to rule-scepticism stem from interposing some mediating device between my understanding of a rule and my application of it, my actual behaviour. Before we can say what rule I follow, we are inclined to think, we must settle on an interpretation of it, just as we must select one empirical hypothesis from the indefinitely many which "fit" the phenomena, as a glove fits a hand or a key fits a lock. But for every interpretation of a rule we can always ask for a rule to interpret the interpretation. This was the problem for the "community-view"--the inclinations of the community do not fix the interpretation of my inclinations, because they can themselves be interpreted in indefinitely many ways. "What this shows," says Wittgenstein,

> is that there is a way of grasping a rule which is *not* an *interpretation*, but which is exhibited in what we call "obeying the rule" and "going against it" in actual cases. (PI §201)

Nothing intervenes between my application of a rule and my understanding of it. The rule does not fit the case in the way that the key fits the lock for precisely the same reasons considered in examining reference and the correspondence-theory of truth. It is the rule itself--*i.e.*, our use of an expression or a paradigm as a rule that determines its cases, just as it is the proposition itself--*i.e.*, our use of *these* signs in *this* way--that determines what fact must obtain if it is to be true. Whether I understand a rule or not is shown by how I use it and explain it. That I understand the rule for "plus" is shown (but not *entailed*) by what I do and by whether what I do accords with past applications. Thus, the possibility of following a rule, Wittgenstein maintains, presupposes an established practice or custom with which one's new uses accord or from which they diverge:

> It is not possible that there should have been only one occasion on which someone obeyed a rule. It is not possible that there should have been only one occasion on which a report was made, an order given or understood; and so on.--To obey a rule, to make a report, to give an order, to play a game of chess, are *customs* (uses, institutions). (PI §199)

What matters here is established practice, not membership in a community, though most of our linguistic practices are communal. Thus, Proto-Robinson Crusoe, abandoned on his desert-island as a small child, could establish practices of his own (assuming he somehow survived) with which he could crudely judge his later behaviour to

accord or fail to accord. But such practices--which might constitute a set of *contingently* private "meanings"[48]--would never approach anything like the complexity of what we normally think of as *language*. Proto-Robinson would have no one to cooperate with, no one to charm, no deals to make, no promises to keep, no salt-shakers to request, no lies to tell, no books or films to discuss, no ideals to criticize or espouse, and so on. The range of purposes and interests that he could have would be sharply limited, deriving only from his interaction with his (regular and predictable!) non-human environment.[49]

V. Two Accounts of Reference

What have the foregoing thoughts to do with Putnam? For one thing, the treatment of Putnam as a sceptic about reference may remind us of Wittgenstein's treatment as a rule-sceptic. And just as Baker and Hacker see Wittgenstein as deriving absurdities from a certain picture of rules, I claim that Putnam is trying to show that a certain picture of truth and reference leads to confusion. Just as Wittgenstein insists on the accord of one's behaviour with a practice as the key to rule-following, so I think that we should treat Putnam's claim that

> In an internalist view also, signs do not intrinsically correspond to objects, independently of how those signs are employed and by whom. But a sign that is actually employed in a particular way by a particular community of users can correspond to particular objects *within the conceptual scheme of those users*. (RTH 52)

Despite Putnam's use of expressions like 'conceptual scheme' and 'community'--and his subsequent remark that "'Objects' do not exist independently of conceptual schemes" (RTH 52)--the emphasis here belongs on "how those signs are employed and by whom." It is because there is "a way of grasping a rule which is *not* an *interpretation*," but rather accord with a custom or practice, that we can refer to the things we think we can refer to. Says Putnam, "[I]t is the practice that fixes the interpretation ..." (RTH 67).

The real weight behind Putnam's complaint with causal theories of reference is that they treat a normative notion as a non-normative explanatory hypothesis. If reference were not normative, then indefinitely many different interpretations could be applied to our actual practice. The spectre of rule-scepticism would loom again: "no course of action could be determined by a rule, because every course of action can be made out to accord with the rule" (PI §201). Finite instances of behaviour can be seen to conform with indefinitely many models or rules for reference, and I have no way of knowing which

---

[48] --Not a logically private language, "which only I myself can understand" (PI §256).

[49] The example is further complicated because the sort of neurological development required for Proto-Robinson to develop even such a limited practice might well not occur without social stimuli. Setting these questions aside--it is not an outright impossibility that someone should have the innate capacity to speak English (*e.g.*), though it is empirically bizarre. See Baker and Hacker (1990).

model is the right one, which rule I do follow. But if I am to be able to use a language at all, then I must be able to intend to use it, and I can intend to do something, only if I have a sense of what would count as success at doing it. "If the technique of the game of chess did not exist, I could not intend to play a game of chess" (PI §337). Since we evidently can use language, rule-scepticism is "*manifestly* self-refuting."[50] Following a rule is a normative, intentional activity, and no adequate account can ignore this.

Reference is, likewise, a normative notion. That is to say, we have standards (rough though they may be), embodied in our linguistic practices, by which we judge to what a person has referred and, indeed, whether she has succeeded in referring at all. If I am to refer successfully to the cat on the mat, then there are some sorts of things that I should typically do (*e.g.*, say, "The cat on the mat," point to the cat, make cat-noises in some contexts ...) and others that I should generally not do (*e.g.*, say, "The cherry on the tree" in non-philosophical contexts, recite from *The Wasteland*, make obscene gestures ...). What counts as referring to the cat varies with context, but in every context there are certain actions that fail to make reference to the cat and others that succeed. Being a capable speaker of English requires that I be able to tell the difference, provided the cases are not too exotic, and that I be able to identify the "strategies of reference" that are most likely to succeed. I can in some contexts succeed in referring to Matthew by calling him "Dan," but I am less likely to be misunderstood if I call him "Matthew."

We can now see how an argument can be retrieved from the ruins of Putnam's model-theory. The correspondence-theorist holds that truth is something to be *explained* in terms of reference and that reference is itself to be explained by a causal theory. Putnam responds that causal theories are non-normative and, so, face the problems of rule-scepticism. The case of reference is just an instance for applying the sort of argument examined in the preceding section.[51]

But as we have seen, there is some variation in accounts of causal theories of reference. Field is primarily interested in the *reduction* of reference to physicalist terms, as though we could find a type-type

---

[50] Baker and Hacker (1984a), p.6.

[51] Baker and Hacker hint at such an application:

> Exactly parallel reasoning would be applicable to the basic contention of truth-conditional semantics, namely that to understand a declarative sentence is to know what would be the case if it were true (and also what would be the case if it were false). Is it thought to follow from this platitude that any form of truth-conditional semantics must be an experimental theory of meaning? Are the truth-conditions of a sentence to be established by bringing it about that it (or what it expresses) is true and then investigating what happens to be true in such circumstances? (*Ibid.*, p.105f)

The absurdity of treating reference as a naturalistic notion (and of explaining truth-as-correspondence in its terms) is the one identified here by Baker and Hacker and the one that Putnam's model-theoretic argument is supposed to reveal.

identity theory for reference. Devitt favours a causal theory "along the lines first suggested by Saul Kripke" (R&T 27), as well as Keith Donnellan and Putnam (in an earlier life). To illustrate the force of the argument that I am attributing to Putnam *via* Wittgenstein, let us look at each of these in greater detail.

On Field's account reference and truth are treated as phenomena on a par with chemical valence. In such a case we want to explain some regularity or similarity among particular instances, and so, we offer hypotheses that would account for the phenomena by seeing them as the consequence of some simpler facet of nature underlying the phenomena. However, once again, this kind of explanation misses the mark, if we attempt to apply it to *normative* "phenomena." To repeat:

> A rule *is not an explanatory hypothesis*, although that a person or a community have such-and-such a rule may be. But to the extent that it is, that 'hypothesis' does not *explain* the behaviour in the sense in which a law of physics explains a natural phenomenon. A rule for the use of an expression (e.g. '+' or 'red') is not a prediction about behaviour, but a standard of correctness. But, of course, that people have such-and-such rules must in general provide grounds for prediction.[52]

To be sure, there are similarities to be found among our uses of a word, and that there are similarities may say something about what kinds of biological organisms we must be and what kind of world we must inhabit. If I could not recognize two distinct occasions of the use of a word as in some way similar, then I could not learn to use the word. But so far this tells us nothing about reference, and neither would going farther in the same direction.

Field supposes that there is something here that we do not know, something that must be explained, if we are to fathom how reference really works. But if there is something we do not know which governs reference, then there are indefinitely many "rules" of reference that we could be following, and we have no way of knowing which one it is that we do follow, or whether we always and all follow the same one. But reference, like the broader category of meaning, is a normative, intentional phenomenon. If I am to be able to refer at all, then I must know at least roughly how I would succeed or fail to do so. This is just what I do not know if reference is like valence.

Of course, Field does not maintain that a theory of the sort he envisions would *fix* reference, merely that it would reduce it. But even without resorting to the arguments I have been offering, it is unclear why reference should not be a "multiply realizable" "phenomenon," just as the functionalists hold that intentional states are multiply realizable[53] and just as no set of necessary and sufficient conditions can tell us what a can-opener is--other than to

---

[52] *Ibid.*, p.92f.

[53] Devitt also suggests this possibility (Devitt, 1981, p.29), but still holds that, "The overall aim in semantics is to explain semantic relationships like designation in nonsemantic terms" (*ibid.*, p.8). Such a functional account, of course, would fall short by excluding the normative. See PP3 225 and Putnam (1988), p.74.

say that it is a device for opening cans. Our *interests* are what make these many realizations *realizations*. There is no infinite disjunction of complex physical relations that just *is* reference, any more than there is some sequential completion of circuits in an electronic chip that just *is* a sorting-programme--though in both cases we can end up treating that particular disjunction or sequence in a particular way. Reference is not the sort of thing about which one needs to give explanatory hypotheses. The only explanation for why 'cats' refers to cats is that *that is how we use the word*. If we used it in some other way, it would not refer to what we now call "cats." But this is not especially deep, and it may be improper to call it an explanation, let alone a "theory." Theories of reference, like theories of meaning generally, are "dead branches in the Tree of Knowledge."[54]

Devitt's approach fares no better. Consider the view that Kripke advances in *Naming and Necessity* in opposition to what he calls the "cluster"-theory of names:[55]

> Someone ... is born; his parents call him by a certain name. They talk about him to their friends. Other people meet him. Through various sorts of talk the name is spread from link to link as if by a chain. A speaker who is on the far end of this chain, who has heard about, say Richard Feynman, in the market place or elsewhere, may be referring to Richard Feynman even though he can't remember from whom he first heard of Feynman or from whom he ever heard of Feynman. He knows that Feynman is a famous physicist. A certain passage of communication reaching ultimately to the man himself does reach the speaker.[56]

This view is plausible and surely tells us something about the transfer of linguistic information--what Devitt calls "reference borrowing."[57] But to the extent that it is correct it is merely a description of one aspect of what I have been calling the interactive account of reference. 'Richard Feynman' refers to Richard Feynman, because that is how we use the name. The notion of a cause certainly plays a role here. The person who mentions the name 'Feynman' to another depends on the laws of physics for the sound-waves created by her voice to impinge on the ears of another, etc. But to call it a "causal theory of reference" is peculiar, much as it would be peculiar to give a causal theory of hockey or a causal theory of dance. If anything, it is an *historical* account of reference, relying on

---

[54] Baker and Hacker (1984a), p.xi.

[55] Devitt follows Kripke in mistakenly attributing the "cluster-theory" to Wittgenstein. See Kripke (1980), p.31; Devitt (1981), p.31.

[56] Kripke (1980), p.91.

[57] Devitt (1981), p.137.

"historical chains"[58] to explain how certain words came to be used as they are. It leaves the "original" fixing of reference *untouched*.

Devitt is aware of this and offers us an account of that "original" fixing of reference, the "initial baptism," as Kripke calls it.[59] But is a "causal" theory any more appropriate here?

> This is connected with the conception of naming as, so to speak, an occult process. Naming appears as a *queer* connexion of a word with an object.--And you really get such a queer connexion when the philosopher tries to bring out *the* relation between name and thing by staring at an object in front of him and repeating a name or even the word "this" innumerable times. For philosophical problems arise when language *goes on holiday*. And *here* we may indeed fancy naming to be some remarkable act of mind, as it were a baptism of an object. (PI §38)

When we speak of x's *causing* y, we typically choose x from a context that also played a role in y's occurrence. If I ask what caused the glass to break, an appropriate answer might be that it was fragile. Another might be that it was dropped on the floor. Which of these is the "correct" answer depends upon my interests in asking the question. But in that case, what is the correct answer to the question of what is the appropriate causal chain that fixes the reference of 'cats'? One alleged causal chain exists between 'cats' and cats or between 'Feynman' and Feynman, but a huge number of other causal chains exist between 'cats' and 'Feynman' and numerous objects, both within and without our customary ontology. As Wittgenstein says of naming, "an ostensive definition can be variously interpreted in *every* case" (PI §28). Why is one of these chains relevant and not the others?[60]

Assuming that we should even speak of isolatable causal chains, the obvious answer, I think, is that one of them is made relevant by our existing linguistic practices. Without those practices that causal chain is of no interest. With some other set of practices another chain, causal or maybe even causal*, could seem relevant:

> When one shows someone the king in chess and says: "This is the king," this does not tell him the use of this piece--unless he already knows the rules of the game up to this last point: the shape of the king. You could imagine his having learnt the rules of the game without ever having been shown an actual piece. The shape of the chessmen corresponds here to the sound or shape of a word. (PI §30)

---

[58] *Ibid.*, p.8n9. Devitt also suggests the adjective "historical" (Devitt, 1981, p.8), but he tends to use "causal." Putnam suggests that neither he nor Kripke was ever sympathetic to the project of reducing reference to causality. See Putnam (1992a), p.221n4.

[59] Kripke (1980), p.78.

[60] See Putnam (1987), pp.37-40, and Putnam (1992a), pp.44-55, as well as Chapter 5, §I below for this kind of argument.

This is not to say, as Wittgenstein clarifies, that one must learn or formulate rules in order to speak: "He might have learnt quite simple board-games first, by watching, and have progressed to more and more complicated ones" (PI §31). But saying "this is the king," explains the use of the king only "if the place is already prepared" (PI §31).

Nor is anything to be gained from inserting a "grounding thought"[61] between the thing designated and the word that designates. For we can just as easily ask why it is a causal link between a cat and a "grounding thought," and not a causal (or causal*) link with some other thing that makes that "grounding thought" refer to the cat. Though one cannot "show" a grounding thought--"a mental representation of [an] object brought about by an act of perception"[62]--to another in order to teach its use, we can yet say that the significance of a perception "would still depend on the circumstances--that is, on what happened before and after" (PI §35) the perception. That we can refer to things at all stems from the fact that things have a *practical* significance for us. Because we are biological organisms with needs and desires, because we can be harmed or benefited, certain pieces of the world *command* our attention, but they do so against the background of our human needs. Indeed, certain portions of the environment command our attention quite actively--*other people*, for example. No perception in itself, no causal chain, *qua* causal chain, between that perception and "the" thing perceived is enough to fix reference.

On the view that I am endorsing the reference of our words depends on the roles that our words play in our language and our activities, as well as on the roles that various objects play in those activities. Moreover, reference is to be understood as an intentional, normative relation, or cluster of relations. At a fundamental level it is *we* who refer to objects in "the world," and we can refer or intend to refer to things, only if there are criteria--open-ended and variable though they may be--for successful and unsuccessful attempts at reference. Sometimes we make mistakes when we try to refer to things. Sometimes we use the wrong word or fail to make ourselves clear. But we often can and do successfully refer to things. When there is real confusion, we usually notice it, and the sort of systematic indeterminacy that non-intentional accounts of reference lead to does not arise because of the roles that words and objects play in our languages and in our activities generally.[63]

That I know what 'cats' means is shown by my ability to use the word in a variety of circumstances--to which there need be no limit--without usually having to be corrected, without my co-linguists always looking puzzled at my talk, and so on. This presupposes that I can use many other words of the language, too, and the same is true for every word. Knowing what 'cats' *refers* to is part of knowing what it means--but not a separable or distinct part. Maybe I need not know how to use 'cats' or the rest of the language in order to know that 'cats' refers to cats, but what I *would* need to know, then, is some other language and how to translate 'cats' into that language. But knowing how to use a great many words entails more than making utterances--I must also be

---

[61] Devitt (1981), p.133.

[62] *Ibid.*

[63] See Hacking (1983), p.105 for a similar point.

able to interact with other people and with objects. That I know that 'cats' refers to cats, 'cherries' to cherries and 'Feynman' to Feynman is shown by the fact, among other facts, that I do not make absurd and unpleasant mistakes in my behaviour toward cats, cherries and the late Richard Feynman. It is shown by the fact that I do not hesitate to accept someone's offer of a piece of cherry pie.

If reference really were indeterminate, it could only be because we had *no guidelines*--or *indefinitely many* incompatible ones--for determining when our attempts to refer were successful. But it is just this sort of case that is entailed by attempting to give a *non-intentional, non-normative* account of reference, in terms, say, of "appropriate causal chains." As long as we cannot single out some one of the many relations between words and the world, there is no way that we can coherently *intend* to refer to anything at all, for indefinitely many objects would be selected by our indefinitely many relations--which is just to say that here there is no such thing as being right or wrong, hence no such thing as intending to refer and no such thing as reference. "[O]nly someone who already knows how to do something with it can significantly ask a name" (PI §31), says Wittgenstein. The mistake here lies in trying to settle upon some non-intentional relation between words and the world.

Chapter 2: Realism and Scepticism

In the preceding chapter I suggested that Putnam's argument against the correspondence-theory of truth and causal theories of reference is best thought of as a Wittgensteinian argument. Putnam's point, I said, is not to show that reference is indeterminate, but that treating reference as an objective, non-intentional relation leads inexorably to such indeterminacy. In the place of causal theories we should embrace an "interactive" conception of reference, according to which the reference of our words is determined by the roles that different words and objects play in our lives.

It is ironic that some commentators criticize Putnam by invoking this intentional and interactive conception of reference. Supposing that Putnam really does think reference to be indeterminate, Blackburn insists, "It is more plausible to suggest that the facts which do determine the interpretation of subsentential components do not lie where Putnam looked for them."[1] And Ian Hacking expresses surprise that Putnam makes mention of "Wittgenstein's argument that meanings cannot be exhaustively given by rules," because, "[t]hat did not mean for Wittgenstein that there was something intrinsically indeterminate and open to reinterpretation in our linguistic *practice*."[2]

It is Putnam's concern in Chapter 1 of *Reason, Truth and History* to argue for the interactive account of reference in order to show why *metaphysical realism* is mistaken. I want now to examine that argument.

I. Brains in a Vat

I noted in Chapter 1 that Devitt was troubled by Putnam's linking of metaphysical, epistemic and semantic issues in his characterization of "metaphysical realism." David Lewis seems to share Devitt's worry. Even if Putnam's argument against the correspondence-theory of truth works, Lewis claims, that is not decisive for realism:

> There would still be a world, and it would not be a figment of our imagination. It could still have many parts, and these parts would fall into classes and relations--too many for comfort, perhaps, but too many is scarcely the same as none. There would still be interpretations, assignments of reference, intended and otherwise. Truth of a theory on a given interpretation would still make sense, and in a non-epistemic way. Truth on all intended interpretations would still make sense.[3]

Lewis' suggestion that the indeterminacy that plagues non-intentional accounts of reference is nothing to worry about is curious. The problem is not simply that there would be indefinitely many interpretations of our sentences that would preserve their truth-values, but that we could not know which interpretation we were making

---

[1] Blackburn (1984a), p.301. See also Benaceraff (1984).

[2] Hacking (1983), p.105.

[3] Lewis (1984), p.231. For related (but different) remarks see Goldman (1986), p.155f.

use of, which "rule" for reference we were following. And if we cannot know this, then it is not just reference and truth that are non-intentional, but our very utterances themselves. For if we cannot know what rules we follow, then we do not follow any rules. In order to follow a rule I must have some conception of what would constitute the success or failure of my attempts--which is just what I cannot have if I can't know what rule I follow. If we cannot tell what interpretation our utterances have, then it is strange to say that they have any interpretation at all.[4] It seems that Lewis has in mind the idea that my words might have meanings inaccessible to me, and that what interpretation they might have is unimportant, as long as the same sentences remain true. But sense can be made of this position only by supposing that meaning and truth are non-intentional, non-normative notions. This is just the supposition that I have rejected.

But Lewis is right to observe that Putnam's Wittgensteinian argument shows nothing about metaphysical realism as a metaphysical (not a semantic) thesis. What link is there between a normative treatment of reference and the thesis that Lewis, Devitt and others want to retain?

Putnam's account of reference comes amidst a discussion of what he calls the "*Turing Test for Reference*" (RTH 9). Suppose, says Putnam, that I am in "conversation" by means of an electronic keyboard with a Turing-machine. The machine is programmed to play the "Imitation Game," to respond eloquently to my questions, appropriately to my remarks and exclamations--so naturally that I think that I am conversing with an intelligent, linguistically competent speaker of English. However, suppose also that

> ... not only does the machine lack electronic eyes and ears, etc., but that there are no provisions in the machine's program, the program for playing the Imitation Game, for incorporating inputs from such sense organs, or for controlling a body. What should we say about such a machine? (RTH 10)

Putnam would have us say that such a machine cannot *refer*, and that its inability to do so arises from its inability to interact with objects in the world, to distinguish between cats and cherries:

> It is true that the machine can discourse beautifully about, say, the scenery in New England. But it could not recognize an apple tree or an apple, a mountain or a cow, a field or a steeple, if it were in front of one. (RTH 10)

That the Turing machine seems to be referring, seems to be "speaking," stems from the fact that we interpret its programme and use it to

---

[4] This is not a verificationist claim, since the meanings of my words are not something that I ordinarily have to verify, though there are instances in which this does happen. Verification cannot be central here for precisely the same reason that non-intentional accounts of meaning do not work. Mistaking this superficial similarity between verificationist claims and claims issuing from a normative conception of meaning is at the root of readings of Wittgenstein which treat his "private-language argument" as worrying about how I could verify the meaning of an act of inner ostensive definition.

accomplish certain ends. But it has no means of relating to the world, other than through its electronic keyboard. Nor has it any *need* to relate to the world. We, however, "are able to perceive, handle, deal with apples and fields. Our talk of apples and fields is intimately connected with our *non-verbal* transactions with apples and fields" (RTH 11). The sounds that we utter are ways of *talking*, because they are embedded in the context of our many other ways of behaving and being in the world. Without that context, arising from a multitude of desires and ends that require us to deal with the world for their satisfaction, there is no language, no meaning and no *reference*.

Now, what has all this to do with metaphysical realism? Putnam's contention is that the metaphysical realist is committed to holding that we *could* refer to things that can play no conceivable role in our projects and activities, that this commitment violates our best understanding of what it means to be able to refer, and hence, that metaphysical realism is false. His argument begins by entertaining the sceptical possibility that, despite our ordinary convictions to the contrary, we might all really just be brains in a vat of nutrient-solution, being fed electrical impulses from some sempiternal computer, which leave us with the impression that we have knowledge of the world. In fact, we are as hopelessly deluded as the Cartesian sceptic imagines us to be in the clutches of the evil demon. Is this situation *really* possible? It violates no constraints of logic or physics (so it seems). Nevertheless, says Putnam, we could not *really* be brains in a vat, because if we were, then the words used to formulate the sceptical possibility would not refer to *brains* or *vats*, but to brains and vats "in the image" (RTH 15), illusory brains and vats--if, indeed, they referred to anything at all. But *ex hypothesi*, we are *not* brains in a vat "in the image"; so, whether we suppose ourselves to be brains in a vat or not, the sentence 'We are brains in a vat' is false, if it means anything.

Let us look more closely at the argument:

1. We speak English if and only if we are not brains in a vat.
2. If we are brains in a vat, then we speak Vat-English or no language at all.
3. Either we are brains in a vat, or we are not brains in a vat.
4. Therefore, either we speak English or we speak Vat-English or no language at all (1,2,3).
5. It is not the case that we speak no language at all.
6. Therefore, either we speak English or Vat-English (4,5).
7. If we speak English, then 'We are brains in a vat' is false (1).
8. If we speak Vat-English, then 'We are brains in a vat' is false, because it means that *we are brains in a vat in the image*, which is not the case if we speak Vat-English.
9. Therefore, 'We are brains in a vat' is false (6,7,8).[5]

It is important to be clear about the import of 'We are brains in a vat'. The possibility that we are brains in a vat is better understood

---

[5] I am indebted here to Bill Barthelemy, though he might not endorse my assessment of Putnam's position. See also Coppock (1987). For a different assessment of Putnam which draws on many of the considerations I outline below see Hymers (1989b). My understanding of Putnam has since changed and, I hope, improved.

as the possibility that we have *always been* and currently are brains in a vat. The significance of this will become clear shortly. As well, although I have formulated the argument in terms of English and Vat-English, it ought obviously to work for other languages; so, we should think of the intent in terms of, say, natural languages and vat-languages, and sentences that can be translated into the English or Vat-English sentence 'We are brains in a vat'. I forego these emendations as obvious and possible.[6]

It should be clear that the argument is valid; so, any complaints must be directed either at the truth of its premises or the strength of its conclusion. Steps 1, 2, 7 and 8 are applications of the interactive conception of reference. 'Brain' refers to brains, 'vat' to vats, 'cats' to cats, 'cherries' to cherries, etc., only if we speak *English, i.e.*, only if we are not brains in a vat, but interact with the world in more or less the ways that we think we do. If we *are* brains in a vat (suspending judgment on the argument's conclusion for the moment), then the things we interact with, and hence, *can refer* to, are not cats and cherries, but illusory cats and cherries--if we interact with anything at all. This is why it is important to insist that 'We are brains in a vat' be taken as implicitly saying that we have *always been* brains in a vat. Were this qualification omitted, we might be speaking English *inside* the vat, at least for a while, due to our prior acquaintance with the language outside the vat.

To say that the interactive account of reference is doing Putnam's philosophical work here is not to say that causal connections are *irrelevant* to reference. It is just to say that reference cannot be analyzed in non-intentional terms. Steps 1, 2, 7 and 8 should also be allowable by any causal theory of reference that is at all plausible.[7] The remaining ones are uncontroversial. Thus, if the argument is vulnerable to criticism, it is at the point of its conclusion.

II. Real Possibility

It is tempting to think, when Putnam concludes that 'We are brains in a vat' is false, that he has proven something, but not what he intended to prove. It may seem that what Putnam's argument shows is that if we were brains in a vat, then we could not *say* or *know* this to be the case. But we could still *be* in a situation that we could not express or know. Thus, Devitt and Sterelny write,

> Putnam wants to show that a certain kind of illusion is impossible. But all that his argument actually shows is that, were we suffering from that illusion, we could not even

---

[6] A language whose speakers had never engaged in such obscure disputes might not have within it the concepts needed to formulate the sceptic's worry, but there is no reason that the concepts could not be imported from another language, given the time and interest.

[7] It is a virtue of Devitt and Sterelny's otherwise weak reading of Putnam's anti-sceptical argument that it recognizes this fact. See Devitt and Sterelny (1987), p.207.

> conjecture that we were. This the realist can, and we do, grant.[8]

On Devitt and Sterelny's reading Putnam tries to move from the conditional claim that if we were brains in a vat, we could not think so, to the conclusion that we could not be brains in a vat. But this simplistic, polemical reading of the argument and their objection in response, take seriously neither the interactive conception of reference, nor their own appeal to a causal theory of reference.

The realist who reasons thus is like the person who thinks that I can experience my own death. It is certainly a logical possibility that I should now be dead, but I cannot coherently conceive of myself *as dead*. In every attempt to do so, I implicitly place myself beyond my own fatality and regard my death as the death of another, who bears my name, past, features, etc. As Sartre says, I doubt my own existence "only in words and abstractly."[9] Similarly, I claim, any attempt to conceive of the possibility that we are and always have been brains in a vat involves an implicit situating of oneself *outside* the vat. In order to say that we could really always have been brains in a vat, but not be able to say so, the metaphysical realist must be able to say *this*. However, it is unclear that being able to say "If we have always been brains in a vat, then we cannot say so," is any more possible than being able to say "We have always been brains in a vat" --if we have always been brains in a vat. If the argument shows that we cannot say the latter, then it shows that we cannot say the former either. When the realist says that we *could* be brains in a vat, she makes a claim that is tacitly counterfactual. Implicit in formulating a description of the possibility is a presupposition that that state of affairs does not obtain. It is as if, in formulating the brains-in-a-vat hypothesis, we entertain it of some other group of people, not of ourselves. It is a possibility for *them*, but never really for *us*.

Let me try to clarify why there is a problem here. The realist cannot purport to hold a position that is genuinely vulnerable to a sceptical worry of the sort represented by the possibility that we have always been brains in a vat and *at the same time* hold that this possibility is *merely* a logical or physical possibility. Putnam holds that the brains-in-a-vat hypothesis violates no logical or physical constraints,[10] but a genuine sceptical hypothesis alleges to be more than merely a logical or physical possibility, *and it is precisely its vulnerability to genuine sceptical hypotheses that has traditionally measured the commitment of metaphysical realism to the existence of a world independent of our abilities to know or describe it.*[11] The sceptic alleges that all the marks of evidence available to us would be explained by our really always having been brains in a vat, and we have no good reason--epistemically speaking--to prefer the hypothesis

---

[8] *Ibid.* I owe my familiarity with this objection to Terry Tomkow. For other versions of it see Brueckner (1986) and Martin (1992), p.8f.

[9] Sartre (1956), p.337.

[10] Hence, there *is* some possible world in which we are brains in a vat, contrary to Michael Kinghan's reading. See Kinghan (1986).

[11] For a similar point see Nagel (1986), p.90.

of our epistemic contact with the real world to that of our hopeless delusion. Each hypothesis, she claims, is as good as the other, when it comes to a general explanation of the variety and nature of our experience. This is the kind of possibility to which the metaphysical realist is committed. The metaphysical realist holds that the world exists independently of our abilities to know about it--*i.e.*, that we might actually be completely mistaken about the way the world is.

It is in response to the sceptical position as I have outlined it that Putnam offers his argument against our being brains in a vat, not in response to the logical or physical possibility that we might be brains in a vat. Putnam wants to show that it is not *really* possible for us to be systematically deluded in this way. But how are we to separate this *real* possibility from logical or physical possibility?[12]

Any satisfactory answer to this question should also be capable of dealing with another complaint that might be raised against the conclusion of Putnam's argument. That conclusion, recall, was that 'We are brains in a vat' is false. On the surface this seems merely to be a claim about what is actually the case, not about what is possible or impossible. Yet Putnam seems to want a stronger conclusion:

> The existence of a 'physically possible world' in which we are brains in a vat (and always were and will be) does not mean that we might really, actually, possibly *be* brains in a vat. (RTH 15)

What could he be getting at here? We can begin to understand "real possibility" by looking at some remarks Putnam makes about Kant:

> [M]y procedure has a close relation to what Kant called a 'transcendental' investigation; for it is an investigation ... of the *preconditions* of reference and hence of thought--preconditions built in to the nature of our minds themselves, though not (as Kant hoped) wholly independent of empirical assumptions. (RTH 16)

I think that we can understand this reference to Kant by considering the status of the synthetic *a priori* in Kant's philosophy.

In *Kant's Theory of Science* Gordon Brittan suggests that one useful way of thinking of the synthetic *a priori* is by appealing to the metaphor of "possible worlds."[13] On Brittan's account, Kant's synthetic truths known *a priori* are propositions that are true in every *really possible world*--they are really necessary truths. Brittan proposes that we think of the set of really possible worlds as a proper subset of the set of logically possible worlds. Thus, while a true synthetic proposition, known *a priori*, is true in all really possible worlds, it is not true in all logically possible worlds--only an analytic or logical truth satisfies that constraint. Moreover, in keeping with the Kantian doctrine that the synthetic *a priori* dictates the form of

---

[12] This crucial sense of "possibility" is often missed. See, *e.g.*, Tichy (1986).

[13] Brittan (1978), pp.20-24. Brittan does not indicate his attitude toward the status of possible worlds, but I shall treat the idea of possible worlds with the *caveat* introduced in Chapter 1, §II.

empirical knowledge, Brittan suggests that a really possible world "is a world that beings like ourselves, endowed with certain perceptual capacities and conceptual abilities, could experience."[14]

Brittan's discussion suggests a promising way of getting at Putnam's position, provided we make some amendments and revisions. First, Brittan treats really possible worlds as though they were none other than physically possible worlds. But since Putnam grants that it is physically possible that we be brains in a vat, I propose that we take the set of really possible worlds to intersect with the set of physically possible worlds. On one hand, there are really possible worlds that are not physically possible, since it is plausible to think that we, or creatures much like ourselves, might have inhabited and known about a world with different physical laws. On the other, it is plausible to think that we might never have existed and that the actual laws of physics would be none the worse for this.

Secondly, Brittan's description of a really possible world is not suitable for introducing the idea, because it includes the modal claim that we *could* have knowledge[15] of such worlds. Let us say, rather, that a really possible world is a world of which we *would* have knowledge, were that world actual.[16] According to these criteria, a world containing nothing but hydrogen-molecules for all time is logically and physically possible, but it is not *really* possible, because it is not a world of which we would have knowledge.[17] Without assuming that an epistemic agent is part of that world we can make no sense of saying that we would have knowledge of that world. But to suppose such an agent is to imagine a different possible world.

Finally, it is important to remember that, although Putnam maintains that there is at least one *a priori* truth,[18] he does not subscribe to the view that the "*preconditions* of reference and hence of thought" are "wholly independent of empirical assumptions" (RTH 16). Rather, for Putnam, concepts are public and couched in language, and we can, therefore, expect that our linguistic capacities will place some constraint on the notion of a really possible world. Indeed, a really possible world will turn out to be one at which we speak a language.

What led me in pursuit of "really possible worlds?" I wanted two things: first, a way of distinguishing real possibility from logical and physical possibility in order to answer the charge that we might be brains in a vat, but not be able to say so, and second, a way of justifying the claim that Putnam's argument has modal consequences. So far, we have partial answers to both questions. Real possibility,

---

[14] Brittan (1978), p.21.

[15] In Kant's philosophy the word 'experience' ('*Erfahrung*') is to be understood as "empirical knowledge." See B 147, 166, 218, 234.

[16] My thanks to Peter Schouls for raising this point.

[17] Treating "real possibility" in this way is akin to denying the characteristic axiom of modal logic S4--*viz.*, ◊◊P --> ◊P. Taking real possibility as possibility *simpliciter*, we could say that worlds that are only logically or physically possible *could have been* possible, but actually are not.

[18] See "There Is at Least One A Priori Truth," PP3 98-114.

based on the account of a really possible world, differs from logical and physical possibility at least in the respect that not all that is logically or physically possible is really possible. Moreover, real possibility has an important link with our epistemic and linguistic capacities, a link which--ostensibly--is absent in the case of logical or physical possibility. As well, since a really possible world is a world of which we would have knowledge, were that world actual, a really possible world is one at which we are not brains in a vat.

But these answers are by no means complete, and they may even provoke some disquiet. For in describing the really possible worlds I have *assumed from the start* that they are worlds of which we would have knowledge. But how does this constitute progress? If this is how we are to understand real possibility, then isn't Putnam's answer to the brains-in-a-vat puzzle *presupposed* by his notion of possibility? Moreover, why should this notion of possibility seem interesting? Surely, it is trivial that in all worlds in which we are not brains in a vat we are not brains in a vat! But what does our assumed knowledge of these other worlds have to do with whether or not we *are* and always have been brains in a vat? Clearly, there is more work to be done.

I suggested earlier that we think of the sceptic's challenge as the presentation of an alternative hypothesis which, if true, would explain the nature and variety of our experience (in a non-Kantian sense of the word). I shall say, then, that a real possibility is one which, if actual, would serve to explain some given phenomenon. Invoking the possible-worlds figure, I can say that a really possible world is a world at which some actual phenomenon is explained by the truth of some hypothesis. That is, what is actually hypothesized to be the case *is* the case at that world, and its being the case is the reason for which the actual phenomenon occurs at that world. Thus, if it is a real possibility that we are and always have been brains in a vat, this alleged possibility can be paraphrased by saying that at some world we *are and always have been* brains in a vat, and our being and having been brains in a vat is what explains the nature and variety of our experience at that world.

However, as it stands, this characterization is too broad. What is needed is a further constraint on the kind of experience that is allegedly explained by our being brains in a vat. Now, it is tempting to say that the Cartesian sceptic raises an alleged *epistemic* possibility in the following sense: it is logically compatible with everything we know that we are right now brains in a vat and always have been. In a sense, this is correct, for the Cartesian sceptic does not--and the metaphysical realist cannot afford to--deny that we have knowledge of our own "subjective" experiences. In its purest form Cartesian scepticism holds, as Kant correctly observed, that "inner experience" is "indubitable" (B 275). However, the sceptic is entitled to treat her challenge as an epistemic possibility of this sort, only if she presupposes that we do not know that we are not brains in a vat. So, this kind of constraint will not quite do.[19]

A better way of putting the sceptic's challenge is to say that our being brains in a vat is logically compatible with our regularly and predictably seeming to say and do the many things that we do seem to say and do in a regular and predictable world. But we must not forget that part of the sceptic's claim is that our so being brains in a vat

---

[19] Thanks to Bruce Hunter. See Chisholm (1989), pp.1-4.

would constitute a viable explanation of our regularly and predictably seeming to do those many things. Without the *proviso* that sceptical error is explanatory of some knowledge that we are conceded to have, the sceptic's hypothesis is not a *real* possibility and, so, does not call into question our justification for assuming the existence of an external world. Real possibility, then, has an "epistemic component," but is not epistemic possibility *simpliciter*. A real possibility is one whose actuality would explain some given knowledge. A really possible world is one at which that knowledge is so explained.

III. Realism and Scepticism: Epistemic Neurosis

Let us return explicitly to the concerns that have implicitly taken us this far. First, what of the modal pretensions of Putnam's argument? We can now see that, if the argument works at all, its implications are not confined to actuality. For we are justified in believing that we have not always been brains in a vat, only if the hypothesis that we have always been brains in a vat need not be taken seriously. That hypothesis need not be taken seriously, only if we have insufficient reason for thinking it really possible. We are justified in believing that we have not always been brains in a vat, only if our having so been would not explain our experience as well as our interacting with the world explains it.[20]

But Putnam's argument is meant to have an even stronger conclusion--*viz.*, that we have *no reason* at all to think that we are brains in a vat and always have been. And we have no reason to think we are brains in a vat, only if the hypothesis that we are would not explain our experience at all. Thus, if we are not and have not always been brains in a vat, then we could not really always have been brains in a vat. If the argument works at all, it has the desired modal consequences.

But could we always have been brains in a vat? Can we just not say so? The realist thinks that we could always have been brains in a vat, but that "the brain-in-the-vat hypothesis, and other such fantasies, are too implausible to take seriously" (R&T 52), that such scepticism is "insoluble" but "uninteresting."[21] But what does it mean to say that an explanatory hypothesis like the sceptic's is implausible? What does it mean to say that it is "unanswerable" (R&T 52)?

To say that the moon is made of green cheese or that we are observed and visited by extraterrestrial life-forms is "implausible." To say that, after a number of years of genuine interaction with the world, I have been kidnapped in my sleep, and my brain has been transferred to a vat of nutrient-solution is also "implausible." Such hypotheses would explain certain phenomena, but there are better explanations. The moon's being made of cheese might explain its brightness in the night-sky, or its pock-marked appearance, but other hypotheses account for these phenomena and for others, too. My kidnapping and envatment would explain my seeming to interact with the world, but my experience suggests that the technology and medical skills needed for such ends do not exist and that no one is that interested in abducting me.

---

[20] One might try to argue that the sceptic's scenario does not cohere well with other well-confirmed scientific beliefs. I shall consider this in Section IV.

[21] Devitt and Sterelny (1987), p.227.

If an explanation is "too implausible to take seriously," then that is because it does not properly account for the phenomena--it fails to predict what it should, or its predictions are inaccurate. But if it does not account for the phenomena, then the worries that it purports to raise are hardly "unanswerable." To say that the sceptical worry is unanswerable might be (a) to concede that it is really possible and that it undermines our justification for our empirical beliefs. If so, it is not too implausible to take seriously. Or it might be (b) to say that it is just logically or physically possible (and so, not worth taking seriously). Such a possibility can never be answered, but only in the sense that it can never be shown *im*possible.

The problem is that the metaphysical realist cannot decide which of these interpretations to embrace, for while she senses something unreasonable about scepticism, her position is dependent on that same unreasonability. The metaphysical realist wants and needs to have it both ways. She must recognize Cartesian scepticism as a real concern, because such a concern amounts to her own claim that the nature and existence of the world are independent of our epistemic capacities. But she must also be able to combine metaphysical realism with a viable theory of knowledge. And that requires either a direct answer to the sceptic (a way of showing an admitted real possibility not to be an *actuality*), or a way of saying *why* scepticism is uninteresting, *why* it is not a *real* possibility. However, the latter option is not open to her, since a reason to hold that scepticism is not interesting is also a reason to hold that the account of independence and objectivity at the heart of metaphysical realism is uninteresting.

But close scrutiny shows that the sceptic's conjecture is not a real possibility at all. Although it violates no law of physics, we could never have a reason for believing it. It is not a rival explanation--although intended as such--because it is self-defeating. It is not "insoluble" or "unanswerable,"[22] because it cannot be true without being false. It is not so much *implausible* as refined *nonsense*, for we cannot coherently doubt that we speak a language. We know that we could not always have been brains in a vat in any situation that preserves our linguistic capacities--in any really possible world.

Here is the link, noted in Section II, between real possibility and our linguistic capacities. A real possibility, were it actual, would explain some accepted knowledge. But any satisfactory account of our experience must include our linguistic abilities. This poses a dilemma. If the sceptic says that our being brains in a vat is a real possibility, then she assumes that we really might not understand our own language--but then she has violated a constraint on her proposal's being a real possibility. If, on the other hand, she chooses to preserve our linguistic understanding, then her claim is false, *ex hypothesi*. Only if we were not language-users could we always have been brains in a vat. But then, it is not clear that 'we' is still an applicable term here. It may be physically possible that there were always brains in a vat and that none of us existed, but this is not a sceptical hypothesis. Confusing it with one is, as Putnam says, "taking physical possibility too seriously" (RTH 15). It may also be logically possible that *we* have always been brains in a vat. But a

---

[22] Remarks such as this distinguish Devitt and Sterelny's views from those which, I noted above, I shall consider in Section IV.

logical or physical possibility is not yet a real one, and it is of no help to the metaphysical realist.

In a sense the worries of the sceptic, as John Wisdom suggests, are like the worries of the obsessive neurotic, but the themes of neurosis are as much exaggerations *of* the norm as deviations *from* it. The neurotic shows us something about our society, our culture, and ourselves; the sceptic shows us something, indeed, even more, about realism. Scepticism lies buried in the "soul" of the realist, and when she tries to distance herself from scepticism by claiming it to be *uninteresting*, she displays a kind of "epistemic neurosis," as I shall henceforth say. That is, the very position that she advocates depends upon a vulnerability to the worries of the sceptic, but those are worries that she cannot afford to take too seriously, if her own positive philosophical programme is not to be undermined. Just as we are reluctant to see ourselves and the ends in which we are implicated portrayed in the behaviour of the neurotic, the realist is reluctant to see realism implicated in scepticism. As Wisdom writes,

> The neurotic, we might say, doesn't believe what he says. Still he does go back at the risk of losing his train to make sure the lights are off. The philosopher doesn't. His acts and feelings are even less in accordance with his words than are the acts and feelings of the neurotic. He, even more than the neurotic and much more than the psychotic, doesn't believe what he says, doesn't doubt when he says he's not sure.[23]

There is a little of the neurotic in us all and more than a little of the sceptic in the realist. Her internal conflict emerges in a second error that Putnam describes: "unconsciously operating with a magical theory of reference, a theory on which certain mental representations necessarily refer to certain external things and kinds of things" (RTH 15). The metaphysical realist wants to express a possibility that in principle cannot be expressed, to say from within the vat, as it were, that we might really always have been brains in a vat. Even while holding that scepticism is unanswerable, the realist supposes that she can reach out and refer to the "real" world. Even while needing to suppose that we really could have always been brains in a vat, the realist clings to the belief that our capacity to refer to things in the world is not genuinely threatened by this. Even while supposing that we really could be utterly deluded, the realist assumes that we are not. How, but by magic, can such assumptions be reconciled?

Maybe magical theories of reference are not *logically* incoherent, but they are *magical*. So, the realist's inner conflict cries out for remedy. She must forsake one of these competing assumptions: (1) that we can talk about the "external" world, or (2) that the objectivity and independence of that world are comprised by a position vulnerable to Cartesian scepticism. It seems evident which is the healthier choice, but the metaphysical realist is reluctant to make this choice, putting in its place a number of defence-mechanisms: "The thorough-going sceptic sets the standards of knowledge (or of rational belief) too high for them ever to be achieved" (R&T 63), says the realist, or, the sceptic changes the meaning of the word 'know'. I am impressed by

---

[23] Wisdom (1957), p.174.

Stroud's arguments that such attempts to evade confrontation with the sceptic do not succeed.[24] But whatever their perceived merits, they should not soothe the anxieties of the metaphysical realist, because to the extent that they are compelling, they are also arguments against metaphysical realism. If the sceptic sets the standards of knowledge too high, then so does the metaphysical realist. But the realist goes even further by supposing that the standards can then also be magically fulfilled. "He, even more than the neurotic ... doesn't believe what he says ..."

Although Putnam's argument focuses on *reference*, it is important to bear in mind that reference is embedded within a context of linguistic meaning. This I argued in Chapter 1. That some relation between a word and an object, or a collection of objects, is a relation of reference stems from the role that that word and the object(s) play in our linguistic and non-linguistic practices. So, to cast doubt on the existence of the external world is to cast doubt on the reference *and* meaning of our words. If we are brains in a vat, then our words--if they mean anything at all--mean something quite different from what they do if we are not brains in a vat. To say that we really could be brains in a vat is to say that we do not know what our words mean. Hence, we do not know what 'We are brains in a vat' means. Putnam's argument is of the following sort: any reason that I could have for doubting that I can have knowledge of the external world is also a reason for doubting that I understand my own words, that my words have any meaning, and hence, a reason for doubting that I understand what is meant by 'We have always been brains in a vat'.[25] Such Cartesian sceptical doubt, as Kant recognized, is self-defeating.[26]

IV. Metaphysical Revisionism

At this point the realist may be tempted to ask just how important scepticism really is to a conception of independence adequate for realism. Traditionally, there is a close link between realism and scepticism. In Kant's characterization the "transcendental realist," on seeing the threat of scepticism, "afterwards plays the part of empirical idealist" (A 369) to avoid that threat. The conception of objectivity and independence embodied in this transcendental realism is such as to say that the world is so independent of our ability to know about it that we might really know nothing of it. Several strategies present themselves for responding to this sceptical worry:

> (i) Argue that the sceptic's account of justification or knowledge is faulty, so that I could be justified in believing in the "external" world, despite the explanatory power of the sceptical hypothesis.

---

[24] Stroud (1984).

[25] *Cf.* Wittgenstein (1972), §383.

[26] See Chapter 6, §III and Kant's "Refutation of Idealism" (B 274-279). His argument, which is largely independent of transcendental idealism--and may even undermine it--is that any reason I could have for doubting the existence of an external world is also a reason for doubting that my representational states are ordered in time.

(ii) Deny that knowledge requires justification at all.

(iii) Argue that although the sceptic's hypothesis is a real possibility, it does not give as good an explanation of intentional phenomena as does the external-world hypothesis.

(iv) Deny that the sceptic's hypothesis is a real possibility.

Of these four the last is unavailable to the metaphysical realist,[27] for it entails rejecting her conception of objectivity. This is the option I have embraced. Is it possible that a metaphysical account of objectivity and independence could be retained without vulnerability to Cartesian scepticism? Strategy (ii) seems counterintuitive and self-defeating, since it removes normativity from knowledge utterly. Most standard uses of 'to know' and its cognates would be undermined, since this account could not distinguish knowledge from *coincidentally* true belief.[28] This leaves the metaphysical realist a choice between (i) and (iii). The latter is *prima facie* more plausible.

Strategy (iii) might be adopted by someone who accepted a quasi-Quinean[29] position of the following sort: the best theory of meaning that we have is a behaviouristic one, and such a theory requires that we have bodies and epistemic contact with the rest of the world. Cartesian scepticism can thus be rejected on grounds that it does not account for the phenomena as well as the real-world hypothesis. However, such a behaviouristic account (it might be maintained) is still an *empirical theory* of meaning, and there are other competing theories, according to which Cartesian scepticism does give as good an explanation as the real-world hypothesis. Hence, scepticism is poorly confirmed, but a real possibility. I have already argued against the central assumption of this view--*viz.*, that meaning is the sort of

---

[27] But it is a variation on the fourth strategy that Descartes offers from the Third Meditation onward. His endorsement of a version of the ontological proof in Meditation IV would entail his holding that--in anachronistic terms--in every possible world we were not systematically deceived about the nature and existence of that world, because such deception is logically incompatible with God's perfection, and God is a necessary being. See CSM II 25ff. His recourse to God's perfection as the key to ruling out the real possibility of the sceptical worries that he raises could be seen to have set the pattern for metaphysical realists in our century, who do not acknowledge the supernatural theory of reference that they implicitly invoke.

[28] This is not to say that no one might be tempted by (ii). Alvin Goldman's earlier views--before process reliabilism, which I shall discuss below--could be seen as an example. See, *e.g.*, Goldman (1984).

[29] I say "quasi"-Quinean, since it is not clear to me that Quine is or would regard himself as a metaphysical realist. But it seems compatible with his views on meaning to make the sort of argument I suggest here, and doing so would make some sense of his claims that "sceptical doubts are scientific doubts" (Quine, 1975, p.68), so that "... in confronting [the sceptic's] challenge, the epistemologist may make free use of all scientific theory" (Quine, 1974, p.2).

thing about which we ought to have an empirical, explanatory theory. So, unless sense can be made of non-intentional accounts of meaning and reference, this option is a non-starter here.

How about the seemingly less plausible strategy (i)? Such a position would be odd. The equation of metaphysical realism with sceptical vulnerability depends on a certain conception of the relation between epistemic justification and the explanatory value of hypotheses--*viz.*, I am justified in claiming knowledge of the world, only if some other hypothesis, *e.g.*, the brains-in-a-vat hypothesis, would not give as good an explanation of my seeming to have abundant evidence for the existence of the world. In order to escape the version of Putnam's argument that I have offered, the metaphysical realist who adopted (i) would have to deny that the availability of competing hypotheses of equal explanatory power is relevant to the justification of my belief in one of those hypotheses.[30] So, given two competing models of, *e.g.*, the structure of the atom, each of which was equally capable of accounting for all the observed and predicted phenomena, and each of which met whatever other standard criteria are desirable in a scientific theory, the *new* metaphysical realist would have to say that one of these theories just was the right one (or more likely to be right). But at this point she begins to sound like the non-intentional theorist of reference, who thinks that causal relations *just are* the ones relevant to the fixing of word-world connections.

Perhaps this dismissal is too quick. I have given little in the way of content to this new, improved version of metaphysical realism. But its proponent might insist that she has some further criterion for distinguishing a justified hypothesis from an unjustified one, a criterion that would allow a rational choice to be made between two hypotheses of comparable explanatory value. Perhaps it might be held that the sceptical hypothesis and the "mundane" hypothesis, although equally explanatory, differ in their degrees of "subjective probability."[31] Subjective probability is, roughly, (a) the actual degree of belief that a person attaches to a proposition, or (b) the degree of belief that she ought to attach to it. In turn, "degree of belief" is to be understood on analogy with betting-behaviour, so that subjective probability becomes the odds that a person would accept, if betting on the truth of a belief, or the odds that she should accept.

Taken as an account of epistemic justification, the first option, as Pollock observes, "has nothing to recommend it."[32] The fact that I *would* choose the external-world hypothesis, or accept a bet about its truth with less favourable odds than those I would insist on for the sceptical hypothesis begs the question about epistemic justification. All that it shows is that I do not believe the sceptic. The concern here is which hypothesis I *ought* to believe; so, the latter account of subjective probability seems more desirable.

The metaphysical revisionist, then, maintains that I ought to accept a bet on the truth of the mundane hypothesis with less favourable odds in preference to a bet on the sceptical hypothesis with better odds.

---

[30] This point is Bruce Hunter's.

[31] I here draw on Pollock's discussion of subjective probability. See Pollock (1986), pp.97-103, 108f.

[32] *Ibid.*, p.108.

However, this leaves unanswered the most important question: *why* should I accept the former sort of bet?[33] The subjective-probability theorist can be taken to be reducing epistemic rationality to prudence, but in what sense would it be more prudent for me to believe in the external world than to believe that I have always been a brain in a vat? Could it be that all of my available *evidence* makes it more prudent? That would beg the question, because what we were looking for was a constraint that could independently tell us what our evidence was evidence for (or what made it evidence, rather than just the appearance of evidence), by giving us a reason to accept one hypothesis over the other. Maybe the point is that it is prudent to believe in the external world, because given what is in fact the case (never mind what evidence or marks of evidence we have), that belief is more likely to be true. However, unless we are justified in believing that there is an external world to which we have epistemic access, how can this assessment be made?

At this point we run up against a picture of epistemic justification quite different from the one implicit in my account so far. It is also quite analogous to the non-intentional accounts of meaning and reference against which I have already argued. The "externalist" holds that epistemic justification is not an intentional relation, such as my being able to give reasons for holding the beliefs that I hold, but an objective, non-intentional relation between knower and known. One such view would allow that whether or not I can make an assessment of prudence regarding my beliefs, I am yet justified in believing that there is an external world, if and only if that belief has a higher subjective probability than a belief in the sceptical hypothesis.

This is not the only "externalist" position that one might hold. "Process reliabilism"--to use Goldman's term--is the view that I am justified in holding a belief if and only if it is produced by a process that is more likely than not to produce true beliefs.[34] It is again unnecessary to justification, on this view, that I have any understanding of the reliable process in question.

What these positions have in common, as I have already said, is that from them one does not see the connection between a belief and its justification as being an intentional one. This separation of belief from justification, or of justification from a person's understanding of that justification, makes it easier for the externalist to draw a sharp line between "*direct* scepticism" and "*iterative* scepticism."[35] Direct scepticism questions our ability to know about the nature and existence of the "external" world, whereas iterative scepticism casts doubt on our abilities to know *that we know* about the world. The externalist might thus maintain[36] that a belief in the external world could be more justified than a belief in the sceptic's hypothesis,

---

[33] Indeed, what if I don't like betting?

[34] See Goldman (1986). I have simplified Goldman's formulation, which refers to "normal worlds," but none of my criticisms turns on ignoring this added complexity. Goldman has himself abandoned the normal-worlds version of reliabilism. See Goldman (1992), p.135-137.

[35] *Ibid.*, p.56.

[36] Goldman does not.

because the appropriate external relation holds between this belief and the world. To hold such a position--that being able to know of the world does not entail being able to know that one knows or is able to --would be to say that it is possible to be a metaphysical realist without conceding that Cartesian scepticism is as important a real possibility as our epistemic contact with the external world, but also without denying that Cartesian scepticism is of comparable explanatory value to the "real-world" hypothesis. Cartesian scepticism (it might be claimed) is a doubt about our second-order knowledge, not about our first-order knowledge. To doubt our ability to single out one "best" explanation for our apparent evidence of the existence and nature of the world is not to doubt our knowledge of the world. All that need be conceded is that we might really have knowledge of the external world without knowing or being able to know that we have such knowledge.

This kind of view suffers from several significant and related difficulties. First, even if it is correct, all that it establishes is the logical possibility of having knowledge of the world without knowing that we have such knowledge. It does not establish that we might really be in that position. Moreover, an argument that showed that we did have knowledge of the world, but no knowledge of that knowledge, would likely give us a reason for thinking that we did have second-order knowledge, for to show that we have knowledge of the world is to acquire the knowledge that we have knowledge of the world. So, it is difficult to imagine what would count as a good reason for thinking that we really might be in this situation.

But there are more serious problems to be faced by the externalist about justification. The consequence of trying to treat intentional notions like reference and meaning as non-intentional, I argued in Chapter 1, is radical indeterminacy and the dissolution of those very notions of reference and meaning. If epistemic justification really is internally connected with the beliefs for which it constitutes justification, then a similar indeterminacy should arise from attempts to reduce justification to some non-intentional phenomenon. That this is the case with reliabilism has been shown by John Pollock.[37]

What, we might ask, would be an example of a reliable process such as Goldman envisions? A plausible candidate might be found in colour-vision, on the assumption that we do have knowledge of the world and that we do so partly as a consequence of our ordinary perceptual processes. Can justification for our beliefs about the colours of middle-sized objects be reduced to the reliability of colour-vision? Notice first that talk of the reliability of colour-vision is much too general, because colour-vision does not tend to produce true beliefs in most of the universe. In the depths of inter-stellar space lighting conditions would be unfavourable for human colour-visual capacities. So, some restriction of context is necessary, if this very plausible candidate for a reliable process is to live up to its promise.

The problem for the process-reliabilist is to provide some non-arbitrary restriction of conditions, while accommodating our customary beliefs about which colour-visual beliefs are justified. For example, it is tempting to say that under white light beliefs produced by the process of colour-vision are justified, because colour-vision is then reliable--*i.e.*, it tends to produce true beliefs about objects viewed

---

[37] In what follows I summarize Pollock's argument with occasional embellishment. See Pollock (1986), pp.118-120.

under such conditions. However, it is equally tempting to say that under blindingly bright white light, or extremely faint white light, beliefs produced by colour-vision are not justified. So, the process-reliabilist must either explicitly exclude these circumstances from her description, C, of the conditions under which colour-vision is reliable or take the counter-intuitive position that even under very bright or very dim white light beliefs about the colours of objects then observed are justified, even though likely to be false.

But there is no end to the number of examples similar to those of very bright light and very dim light. Suppose that I have taken LSD and have to make judgments about the colours of objects in my nearby surroundings, illuminated by white light. The internalist can discern two kinds of cases here, the one in which I know that I have taken an hallucinogen, and the one in which I do not. And she can go on to say, *e.g.*, that in the former case my beliefs about the colours of the objects around me are unjustified, but in the latter they are justified or justifiable. The reliabilist, however, must treat both cases exclusively as ones of justified or unjustified belief.[38] And as before, she must either explicitly exclude such circumstances in her description of C, or accept the odd claim that my beliefs would be justified, even though very likely to be false.

The problem is that with every such defeating case that she rules out the reliabilist comes closer to saying that a belief is justified if and only if it is true. The reliabilist wants to be able to say that colour-vision is reliable under *normal* conditions. But normal conditions do not cry out for identification any more than the class of cats cries out for identification. Normal conditions are normal relative to our interests and intentions in so classifying them. The "normal" in this context is normative.

We can put the point another way. Rather than as trying to identify the cases in which a given process is reliable, the reliabilist might be seen as trying to identify the process that is reliable. For example, is it colour-vision, or colour-vision-on-Earth, or colour-vision-on-Earth-under-white-light, or colour-vision-on-Earth-under-(not too bright or too dim)-white-light ... ? The problem now is that no process cries out for identification. No non-intentional process *just is* the "justification-process," any more than a non-intentional connection between words and things just is "the" reference-relation. There are indefinitely many different processes, or indefinitely many different true descriptions of context for a given process, and the reliabilist cannot select one of these without either implicit appeal to intentional notions like "normal conditions" or the unpalatable concession that a belief is justified if and only if it is true.[39]

---

[38] My belief that I have taken LSD might call into question the justification of my beliefs caused by colour-vision in these circumstances, but only if it is caused by a reliable process. That my belief that I have taken LSD is caused any more reliably under such conditions than my colour-visual beliefs is by no means clear.

[39] Goldman has tried to rebut Pollock's criticism by claiming that reliabilism offers a *description* of our actual pre-reflective intuitions about epistemic justification, not an analysis of the normative in terms of reliability. He writes of Pollock's objection:

An option remains for the metaphysical realist. She might try to give an account of the independence of the existence and nature of the world from our epistemic capacities, which does not acknowledge the real possibility of Cartesian scepticism. This view might consist in accepting a vulnerability to an even more radical kind of scepticism than the Cartesian proposes.[40] The metaphysical realist might try to maintain that we could be mistaken in *all* of our beliefs, not just our beliefs about the "external" world. This proposal meets with trouble almost immediately. What sense of possibility is relevant here? It cannot be epistemic possibility, since this would be question-begging: that we are mistaken in all of our beliefs is compatible with everything we know, only if we are assumed to know nothing. It cannot be real possibility, because that kind of possibility presupposes knowledge of something. And as was the case with Cartesian scepticism,

> It would be a mistake to suppose that ordinary epistemic evaluators are sensitive to these issues. It is likely--or at least plausible--that our ordinary apprehension of the intellectual virtues is rough, unsystematic, and insensitive to any theoretical desirability of relativization to domain or environment. Thus as long as we are engaged in a description of our epistemic folkways, it is no criticism of the account that it fails to explain what domain or environment is to be used. (Goldman, 1992, p.162f)

Such a retreat would greatly increase the plausibility of Goldman's account, for reliability undoubtedly has something to do with our sense of what judgments are epistemically justified and what ones are not. (I suspect that internalism also plays a role in our intuitions--though I can offer no proper empirical evidence. But inasmuch as the very notion of justification is at its heart an internalist one, it is unlikely that internalist intuitions would be absent. It is in virtue of its ability to pick out cases marked "justified" by internalist norms that reliability seems related to justification at all.)

Unfortunately, despite this apparent solution to the problem raised by Pollock, Goldman still believes that epistemology has a normative task to fulfil in addition to its descriptive task, and he persists in claiming that "normative scientific epistemology should follow in the footsteps of folk practice and use reliability (and other truth-linked standards) as a basis for epistemic evaluation" (*ibid.*, p.164). And he also seems to hold that it might be advisable for a "scientific epistemology" to abandon the idea that one need be able to become aware of the process that produces a belief in order to be justified in holding that belief. If what were meant here were "causal process" --in contrast with norms--then there would not need be any conflict between Goldman's position and my own. However, in keeping with his conflation of the normative and the empirical, Goldman's "processes" seem to include rules, methods, procedures and the like--*i.e.*, normative phenomena. "One example is reasoning processes, where the inputs include antecedent beliefs and entertained hypotheses" (*ibid.*, p.115). As long as these elements of his view remain, Goldman's position is vulnerable to Pollock's argument or ones parallel to it.

---

[40] Thanks again to Bruce Hunter.

the logical or physical possibility of total error is insufficient to distinguish metaphysical realism from internal realism.

One way of expressing the sort of possibility in question might be as follows: it is logically compatible with our having all the beliefs that we actually have that all those beliefs should be false. This is more than just the logical possibility of "total error," because it specifies that we would still have the same beliefs that we actually have. It is not just some set of beliefs that is supposed to be possibly false, but our actual beliefs. However, it is not presupposed in this possibility that we know *what* beliefs those actually are. The case might be further detailed by extending it from our actual beliefs to our actual epistemic capacities. The unrepentant metaphysical realist might insist that given whatever epistemic capacities we actually have (but not the assumption that we know what those capacities are), for every set of beliefs, S, produced by those capacities, C, there is a world, W, at which we have C and S and at which all the members of S are false.

However, acknowledging such a sceptical possibility displays a kind of psycho-philosophical disorder comparable to the traditional metaphysical realist's half-hearted attempt to acknowledge Cartesian scepticism. For if I (posing as a metaphysical realist) am to accept a genuine vulnerability of my position to such scepticism--and this is the point of the specification that my actual beliefs might all be false, given my actual epistemic capacities--then I have to accept that my sceptical worry is vulnerable to the same doubt. Any reason that I could have for doubting all of my beliefs in this way would be a reason for doubting my doubt. To the extent that this scenario is a possibility of the sort that the metaphysical realist requires, it is, paradoxically, not such a possibility. Without a non-intentional account of meaning and truth, this sort of scepticism falls to the same sort of argument that showed Cartesian scepticism to be self-defeating. For it could not be *true* that all of our beliefs are false, unless truth were something other than just a property of beliefs and statements--*i.e.*, unless truth were "radically non-epistemic,"[41] a property that could be possessed by so-called "propositions" apart from the real possibility of their being believed or asserted.

This attempt to rescue metaphysical realism is a more general version of a strategy that we encountered in Chapter 1. While arguing against rule-scepticism, I claimed that we could not merely seem to speak a language--a premise that is also part of the framework for my arguments in this chapter. It is easy to imagine a metaphysical realist insisting that we really could merely seem to be language-users, that our words and expressions really might be utterly devoid of meaning, or possessed of some meaning of which we are completely ignorant. My response now, as it was then, is to say that we can seem to be using words correctly or incorrectly, only if we can recognize some standard relative to which we seem successful or unsuccessful. Otherwise, this alleged "seeming" can be no more than a brute sensation, like a pain or an itch. (And that just means that here we can't talk about "seeming.") On the other hand, if there is some standard, relative to which we seem on particular occasions to be successful or unsuccessful in our apparent use of words, then we

---

[41] Putnam (1978), p.125. I shall return to the question of truth as a concept with an epistemic component below.

really do use words and speak a language. Understanding an expression just is knowing how to use it in particular cases--*i.e.*, knowing what counts as correct or incorrect use, relative to a standard of use.

V. Realism? Anti-realism? Historicism?

These considerations suggest that the metaphysical realist is unlikely to find a plausible account of her position that does not depend on treating Cartesian scepticism as a real possibility, which purports to give as good an explanation of the nature and variety of our experience as would our having knowledge of the "external" world. That real possibility gives content to the notion of objectivity and independence on which metaphysical realism rests. But is it possible that a form of realism might be embraced that settled for a less extreme account of independence? Let us consider what this might mean.

"Not empiricism and yet realism in philosophy," says Wittgenstein, "that is the hardest thing."[42] Does abandoning "metaphysical realism" mean that we give up all title to "realism?" I suspect that this is largely a matter of terminological preference. In rejecting the sceptic's account of independence, I have rejected the view that the world exists independently of our actual abilities to know about it. Similarly, I have rejected the view that the world is independent of our possible abilities to know about it, since it is *really* impossible that the sceptic is correct. But I have not rejected the idea that the world *could* have existed independently of our actual or possible abilities to know about it. Nor have I rejected the view that the world exists independently of our *actual knowledge*, or, indeed, of ourselves: the human race and every other "intelligent" form of life in the universe might cease to exist, but the world would continue without us, just as it existed before anyone lived to know of it. On a less cosmic level, many particular beliefs that we hold may well turn out to be false, just as many of our past beliefs have so turned out. What is undermined by showing that we could not really be mistaken about all our beliefs is not this kind of independence, nor the "pragmatic" or "internal" realism--the realism with a "small 'r'"--that Putnam advocates, but a seductive metaphysical conception of independence and objectivity that sometimes pretends that there could be no alternative to realism but to suppose that the world "is constituted" by the mind in some way,[43] or that, in Lorraine Code's words, "A knower must either value objectivity absolutely or succumb to the vagaries of subjectivity run wild...."[44] The metaphysical realist tends to treat attempts at rejecting such an exaggerated notion of independence as sliding automatically into idealism or relativism, obscuring the possibility of a middle ground by speaking broadly of "mind-independence" or independence from "the mental."

Dummett's portrayal of realism and anti-realism, mentioned

---

[42] Wittgenstein (1983), p.325.

[43] Discussing a paper at the University of Alberta (Sterelny, 1991), Kim Sterelny seemed to advocate this false dichotomy.

[44] Code (1991), p.30. Code's "epistemological relativism"--in spite of its nomenclature--has much in common with my position.

earlier,[45] plays into the hands of the realist. Not only does he conflate semantic, epistemological and metaphysical issues, but he conflates the wrong ones, as it were. He equates the possession of "an objective truth-value, independently of our means of knowing it" with a statement's being "true or false in virtue of a reality existing independently of us."[46] But reality can be independent of us and not be independent of our ability to know the truth-values of sentences.[47]

The dichotomy of the objective and the subjective is a false one, but "the dissolution of a dichotomy does not render its terms meaningless. Rather, it denies both terms the absolute force that the oppositional structure of the dichotomy confers."[48] Objectivity and subjectivity are directions on a continuum, not utterly independent opposites, complete unto themselves, and as I shall argue in later chapters, this non-dichotomous rendering of terms has important consequences, not only for our knowledge of the world, but for our knowledge of other minds and cultures and of ourselves.

Now, it is tempting to say that if the existence and nature of the world are not independent of our abilities to know about them, then they must be dependent *on* our epistemic abilities. This makes it sound as though the nature of the world is causally dependent on us, as though our minds "make up" the world. We can quite validly resist the temptation to slip into such idealism. The existence and nature of the world do not depend on *us*, or our *knowledge*, but on our epistemic capacities. And it is not a *causal* dependence, nor anything even like a causal dependence. Rather, the dependence consists in this: that in every world in which we speak a language--every really possible world --the sceptic's hypothesis is false. In all such worlds we can have knowledge.[49] Furthermore, the dependence goes both ways, since every world in which we can and do have knowledge is a really possible world, and every really possible world is a world in which we can and do have knowledge. So our epistemic capacities depend on the existence and nature of the world. This is hardly surprising, since we are part of the world, since we are organisms in a complex eco-sphere,

---

[45] See Chapter 1, §I.

[46] Dummett (1979), p.146.

[47] Dummett has written much since this characterization originally appeared, but I have elected not to test those waters here.

For another conflation of these two varieties of independence see Lenin's criticisms of Neo-Kantianism and related doctrines. He writes:

> In fact the doctrine of the independence of the outer world from consciousness is the fundamental proposition of materialism. The assertion that the earth existed before the appearance of man is an objective truth. (Lenin, 1927, p.96)

[48] Code (1991), p.30.

[49] To return to modal logic, rejecting the characteristic axiom of S4 would let us say that the dependence of the world on our epistemic capacities is a *logical* one, since in every (really) possible world it is both the case that the world exists and we can know of it. However, I shall refrain from this use of the word 'logical'.

conjunctions of physical processes in a universe of physical processes. Our epistemic capacities are always found together with a knowable world--or would be, since we do not have the option of abandoning the one actual world. Neither dependence-relation is a cause for panic, let alone idealism. Indeed, put in perspective, they seem *almost*, but--I hope--not quite, uninteresting.

It is a matter of little intrinsic importance whether the position I am advocating here is thought of as a kind of realism or anti-realism, as long as it is clear what the position is. Since it is not a form of idealism, it is tempting to call it "realism," qualified by some appropriate adjective, maybe. However, my position might also be considered a variety of historicism, the view that our historical and cultural circumstances limit our knowledge and constrain the terms in which we think of ourselves and our surroundings. That this is so may not be immediately apparent, but there are two interesting aspects of epistemic constraint embodied in this view. First, in rejecting metaphysical realism I accept a dimension of epistemic constraint. The metaphysical realist, while needing to take vulnerability to Cartesian scepticism seriously, also acts as though such vulnerability were not genuine. Thus, metaphysical realism is committed to the supposition that we *can* have knowledge of a world whose nature and existence are utterly independent of our epistemic capacities, though how this is possible, as we have seen, is unclear. By contrast my position aspires to more modest, and more coherent, goals by supposing that we can have knowledge of a world that is independent of our descriptions of it, but not of our epistemic capacities. If such a view *seems* more extravagant, it does so only from the perspective of the Cartesian sceptic, which the traditional realist conveniently adopts whenever something less extravagant than metaphysical realism is proposed.

But there is a more interesting constraint on our knowledge lurking here, and it is a direct consequence of adopting an interactive conception of reference and treating meaning as *use*. It is a central feature of this conception of meaning that the significance of a word or phrase is given by its typical context of utterance. We learn a use by learning to recognize relevant similarities of context from one occasion to the next. However, it is an equally central feature of this account of meaning that meanings can and do change, for the simple reason that the way in which a word is used can and does change. No two occasions of use are utterly qualitatively identical; each context differs from every other context, at least by being spatio-temporally distinct,[50] and, less abstractly, in features to which our attention can be drawn. We single out features that strike us as similar and so delineate a loose set of contexts, which comprises the standard or literal meaning of a word or phrase. But on any particular occasion *some other* contextual feature may strike us, or in a very "different" context we may perceive an unnoticed or under-emphasized similarity. Such is the case with, *e.g.*, metaphor.

What this shows, I think, is that what we can mean or intend to say is constrained by the contexts of utterance with which we are familiar. This is not an absolute constraint, however, because we are capable of noticing new similarities. The uses of words can change

---

[50] There are no sharp borders for contexts, but each has a focus, much as Leibniz's monads represent the universe from a point of view. For an interesting discussion of context see Derrida (1988), pp.1-23.

over time as new similarities of context dominate or replace old similarities. But it remains the case that in order to speak intelligibly we depend on a tradition or custom of linguistic use. This shows that what we can mean--and what we can know--are constrained in an interesting and relevant way by historical and socio-cultural contexts. Again, the tension between tradition and novelty is in place. Custom is not an absolute constraint, but it is a constraint nonetheless, one which is exacerbated by separations of time and place and, therefore, of language.

Now, it is also worth observing that the view I am arguing for would probably be classified by Dummett as a kind of "anti-realism." This is because, not only do I reject metaphysical realism, but despite an unclarity that I have already noted in Dummett's early description of anti-realism, my position fits the portion of that characterization pertaining to semantics and our epistemic capacities. Sentences do not have truth-values independently of our abilities to know those truth-values. That is not to say that sentences do not have truth-values independently of whether or not we know them. A sentence is not true, or false, simply because we regard it as being true, or false. We can be mistaken--just not *all* or a *majority* of the time. For a sentence to be true or false, it must be the case that we really could make some determination in this regard--there must be evidence or reasoning or experiences that would count for or against holding it true.[51]

Such a conception of truth is apt to sound like a version of verificationism, but it differs importantly in two ways. First, it involves no commitment to a verification-theory of meaning. I am not saying that meaning is given simply by the method(s) that would verify or falsify declarative sentences. Such methods are not irrelevant, but meaning is in the first instance a matter of use, and it is from this central feature that any relevance for verification or falsification derives. If verification- or falsification-conditions do tell us something about the meaning of a sentence, then that is because of the way in which the sentence is used--*e.g.*, in an experimental setting. Not only is the traditional verification-theory of meaning self-defeating,[52] since its very formulation is unverifiable, but it also presupposes that the individual declarative sentence is the unit of significance. Such presupposition is cast into doubt, on one hand, by Quine's arguments that "our statements about the external world face the tribunal of sense experience not individually but only as a corporate body,"[53] and on the other, by recognizing that declarative sentences are just one kind of speech and writing among many. The supposition that declarative sentences are *basic* stems in part from a very specific and limited picture of language, a picture in which language is most importantly used when it states *facts* and the *results* of inquiry--a scientistic and authority-confirming picture. As well, though verification is not irrelevant to the account of truth I endorse, it is important not to read another scientistic feature into that account. There is no over-arching method by means of which we can

---

[51] This clarifies why Devitt's insistence on separating semantic and metaphysical issues is somewhat tendentious. See Chapter 1, §I.

[52] See, *e.g.*, RTH 105-113 for an argument supporting this claim.

[53] Quine (1980), p.41.

"verify" sentences about diverse subject-matters. Whether and how a sentence is to be deemed rationally acceptable varies according to the kind of sentence it is. Procedures for making judgments about the mechanics of falling bodies differ greatly from those relevant to assessing the quality of a meal or giving a verdict in a court of law.

Now, Putnam sometimes tries to capture the conception of truth that I am advocating by saying that "truth is an *idealization* of rational acceptability" (RTH 55). Is there a sensible way of understanding this remark? There are many things that we might be tempted to read into Putnam's claim. In particular, his intention in speaking of an "idealization" of rational acceptability is difficult to pin down. Consider his initial explication:

> We speak as if there were such things as epistemically ideal conditions, and we call a statement 'true' if it would be justified under such conditions. 'Epistemically ideal conditions', of course, are like 'frictionless planes': we cannot really attain epistemically ideal conditions, or even be absolutely certain that we have come sufficiently close to them. But frictionless planes cannot really be attained either, and yet talk of frictionless planes has 'cash value' because we can approximate them to a very high degree of approximation. (RTH 55)

The analogy with frictionless planes is doubly unclear (and, I shall suggest, misleading). First, whether or not a plane is frictionless or nearly so is not supposed to be a matter of who observes it. But, as Putnam concedes, "rational acceptability is both tensed and relative to a person" (RTH 55). Is ideal rational acceptability not similarly relative? One is also tempted to ask, how can we ever adequately specify such ideal conditions? No list of necessary and sufficient conditions has been or could be given. Even if a list could be given, for whom would the conditions be ideal? And surely we cannot make sense of the idea of being in a position to evaluate every sentence.[54]

The second way in which the frictionless-planes analogy is unclear has to do with the question of approximation. The better our warrant for a statement, it might be plausibly maintained, the more likely it is (from our evidence-relative perspective) to be true. But does this mean that we could--in "theory" if not in practice--have perfect justification for asserting a statement? Does it mean that we could have *conclusive* reasons for asserting or believing the statement? In the "Introduction" to *Realism and Reason* Putnam throws cold water on this suggestion, contrasting his view with the view that "ordinary-language-sentences about material objects outside of theoretical science could be conclusively verified" (PP3 xvii). But if idealized justification does not consist in having conclusive reasons, then it seems as though an ideally justified statement could still be *false*.[55]

How should we deal with these two sets of worries? We can begin to accommodate the first set by recognizing that some of them place demands on Putnam that presuppose him to be offering a systematic *theory* or *definition* of truth. Unfortunately, this presupposition is

---

[54] See Field (1982). Thanks to Bruce Hunter for similar concerns.

[55] See Goldman (1986), p.146.

occasionally given credibility by Putnam himself, who, at one point, suggests another analogy between "truth" and "electrical current." Criticizing disquotational accounts of truth, he writes ...

> ... describing assertibility conditions for 'This sentence is true' ... does not preempt the question 'What is the nature of truth?'... . If a philosopher says that *truth* is different from *electricity* in precisely this way: that there is room for a theory of electricity but *no room* for a theory of truth, that knowing the assertibility conditions is *knowing all there is to know* about truth, then ... he is denying that there is a *property* of truth (or a property of rightness, or correctness), not just in the realist sense, but in *any* sense. (PP3 xv)

Despite this comparison--which Rorty credits to Putnam's inability to renounce fully his metaphysical realist ways[56]--we should be wary of the presupposition that it might be seen to support, for Putnam says that he is "not trying to give a formal *definition* of truth, but an informal elucidation of the notion" (RTH 56). He has gone on to write,

> Now, the picture I have just sketched *is* only a "picture." If I were to claim it is a *theory* I should be called upon at least to sketch a theory of idealized warrant; and I don't think we can even sketch a theory of actual warrant ... let alone a theory of idealized warrant.[57]

The point is not to give an exhaustive description of some set of conditions under which each and every sentence could be verified. The idea that truth is something about which we ought to have a theory is just the idea that we should be led away from by the considerations of this chapter. Truth, like meaning, is a normative notion. That we understand truth is shown by our ability to speak a language--to make statements, to ask questions, to expect and understand answers, to hope we are not mistaken, etc.

Truth is not a very deep notion, but neither is it a dispensable one. Nothing hidden awaits our discovery in some science of the truth in the way that sub-atomic particles and quasars remained unheard of by human beings for millennia. If something were hidden here, then we could be completely mistaken about the meaning of 'truth'. But the meaning of 'truth' just is the way or related ways in which 'truth' is used, and if we speak English, then we cannot be completely mistaken about that--though we may have some difficulty in giving a clear overview of how the word is used. There is nothing here that wants explaining--except in the sense in which we explain the meanings of words--and so, no need for an empirical theory. The use of 'truth' is connected with the use of 'warrant' and 'justification' and, as Putnam says, recognizing that a warranted assertion can lose its warrant is part of what "show[s] that one has the concept of warrant."[58]

---

[56] See Rorty (1992a).

[57] Putnam (1990), p.42.

[58] *Ibid.*, p.22.

Still one might wonder whether even a *picture* can be sketched of idealized justification that is clear enough to account for the person-relative nature of justification. Is there some set of conditions under which *everyone* would be justified in holding some arbitrary sentence justified?

There is a simple way of spelling out the idea of truth as idealized rational acceptability. It is a picture that can be pieced together from Putnam's writings and the arguments I have been presenting in these first two chapters. I shall now assemble some of those pieces.

VI. Truth

I noted that the comparison of epistemically ideal conditions to frictionless planes was not only unclear, but misleading. Putnam himself seems concerned that we not take the analogy too literally:

> The simile of frictionless planes aside, the two key ideas of the idealization theory [?!] of truth are (1) that truth is independent of justification here and now, but not independent of *all* justification. To claim a statement is true is to claim that it could be justified. (2) Truth is expected to be stable or convergent; if both a statement and its negation could be 'justified', even if conditions were as ideal as one could hope to make them, there is no sense in thinking of the statement as *having* a truth-value. (RTH 56)

These remarks are related to points that Putnam makes elsewhere. Truth, he says, is a "property," or "a 'substantial' notion" (PP3 278), "a substantive" or "substantial property of assertions" (PP3 280f), or "some kind of correctness which is substantial and not merely 'disquotational'" (PP3 246) or again a "*normative* property."[59] This last remark he elaborates in response to criticism from Rorty:

> To say ... that truth is a *normative* property is to emphasize that calling statements true and false is *evaluating* them; and evaluation presupposes standards, among them the laws of logic. Our standards of truth are extendable and reformable; they are not a collection of algorithms. But for all that, there are statements that meet them and statements that do not; and that is what makes truth a "substantial" notion.[60]

These various comments afford a number of pieces that fit into the picture of truth as idealized rational acceptability. First, truth is an *epistemic* notion and so, a *normative* one, since knowing is linked with justification, and justification is intentional and normative (*contra* externalism). Secondly, the normativity of truth cannot lie simply in "justification here and now" or in its being "a word we use to pay 'compliments' to sentences"[61]--the equivalent of "*assenting* to

---

[59] Putnam (1988), p.69.

[60] Putnam (1992b), p.436.

[61] *Ibid.*, p.437.

the statement" (PP3 245) held true. There is a difference between holding a sentence true and its being true--truth is "substantial."

That truth is an epistemic notion is the most important theme of truth-as-idealized-rational-acceptability. Even if no sense can be made of the role of idealization in this formula, no philosophical aspersions are thereby cast upon the claim that truth and warrant are internally related. In learning the use of one concept I learn the use of the other. This internal relation is a simple consequence of the assumptions underlying the argument that showed that 'We are brains in a vat and always have been' must really be false. For the examination of the notion of a really possible world, remember, showed that really possible worlds are the worlds in which we would have knowledge, and they are also the worlds in which we would speak a language. For truth to be *non-epistemic*, there would have to be statements whose truth-values were determinate but could not really be determined. But there is no really possible world at which a statement is made and at which it is not really possible to decide whether it be true or false (or neither). Nothing is true (or false) at worlds in which we could have no knowledge, for the simple reason that truth is a property of statements (and of beliefs). Putnam responds again to Rorty:

> [W]hile it is true that the sky would still have been blue (indeed, bluer!) even if language-users had not evolved, it is not true that *true sentences* would still have existed. If language-users had not evolved, there would still have been a world, but there would not have been any *truths* about the world.[62]

So, though it is (*actually*) true that there might have been only hydrogen-molecules existing all through time, it would not have been *true* that there were only hydrogen molecules. Truth is epistemic, because knowledge is of the true, *and* there is (and could have been) no truth that we could not really (have) know(n) or expressed. Since truth does not transcend all epistemic or linguistic practices, and since both these sorts of practices are *normative*, truth is also a normative notion.[63] If a statement is true, then it can, or could be,

---

[62] *Ibid.*, p.433. Putnam's second sentence states the case too strongly here, since we might want to hold, as I shall in Chapter 8, that animals incapable of language might still be capable of beliefs. Such beliefs--in order to be beliefs--would have to be capable of possessing truth-values. However, it is quite compatible with the account of truth and possibility that I present in this chapter to hold that if there had been no language-users, but other animals capable of a certain degree of normative behaviour (*e.g.*, dogs can herd sheep) it would still be *really possible* for the beliefs of those animals to be expressed, even though they would never *be* expressed.

[63] Pollock writes:

> We do not need the concept of truth in order to affirm a thought while thinking that thought. The affirmation is part of the thinking. But in order to affirm a thought while thinking about it, we do need the concept of truth. The ability to ascend a level and think about our thoughts is

justified. Its being justified here and now does not entail its truth, but its truth entails the real possibility of its being justified.

But what of "idealization" in the picture of truth as idealized justification? I suggested that this component is not of decisive importance to my position or to Putnam's, despite his use of the expression. He sometimes seems in clear accord with this remark:

> "[T]ruth is idealized rational acceptability." This formulation was taken by many as meaning that "rational acceptability" ... is supposed (by me) to be *more basic* than "truth"; that I was offering a *reduction* of truth *to* epistemic notions. Nothing was farther from my intention. The suggestion is simply that truth and rational acceptability are *interdependent* notions.[64]

Nonetheless, it is tempting to inquire after the intuitions that led Putnam to paint his picture in terms of an *idealization* of warrant. I want to suggest that the sort of idealization in question can be extracted directly from the presuppositions of the argument that 'We have always been brains in a vat' must really be false. Putnam does not propose this account, but I think that it captures what he seems to have had in mind in offering the idealization-"theory" of truth.

Here is the solution I offer: to say that a proposition is ideally justified is to say that it cannot be coherently doubted. 'We have not always been brains in a vat', *e.g.*, cannot be coherently doubted, and so, is ideally justified and *a fortiori* true. My proposal makes good sense in a number of ways: (1) it flows naturally from the rejection of metaphysical realism and the acceptance of an account of linguistic meaning as use; (2) therefore, it is directly connected with the contention that truth is an epistemic and normative notion; (3) it accommodates the claim that for evaluating most propositions "we cannot really attain epistemically ideal conditions" (RTH 55), but we

---

[63] required for the operation of defeasible reasoning, and that in turn requires that we have the evaluative concepts of truth and falsity. (Pollock, 1986, p.165)

Pollock regards truth as normative or "evaluative." He himself draws attention to a similarity between his views and "the somewhat cruder" (*ibid.*, p.148n16) ones of Putnam and Dummett. But I would be reluctant to accept the firm distinction between levels of language. If we did not have the "language-game" of "thinking about our thoughts" and assessing them, it's not clear to me that we could have the language-game of simply thinking our thoughts. This is not to say that we could not then have beliefs and other intentional phenomena, since believing something is not a matter of "thinking a thought" or expressing the "thought" in some other way and certainly not of "manipulating sentences in the language of thought" (*ibid.*, p.163). But these points go well beyond what I can effectively argue here. Suffice it to say that Pollock does not accept an "interactive" account of reference, such as the one I endorsed in Chapter 1, and he is an individualist when it comes to the classification of intentional phenomena. See Chapter 7 for a discussion of individualism and anti-individualism.

[64] Putnam (1988), p.115.

can still approximate them; (4) it avoids the complaint that a statement might be ideally justified and still be false--idealized justification gives conclusive reasons.

The virtues of this approach can be brought out by applying it to one of Putnam's examples. In *Realism and Reason* he contrasts his position with the view that statements "outside of theoretical science" can be conclusively verified:

> Consider the sentence 'There is a chair in my office right now.' Under sufficiently good epistemic conditions any normal person could verify this, where sufficiently good epistemic conditions might, for example, consist in one's having good vision, being in my office now with the light on, not having taken a hallucinogenic agent, etc. (PP3 xvii)[65]

Putnam speaks here of "sufficiently good epistemic conditions," rather than of *ideal* ones. We could treat this caution in two likely ways: (i) Putnam is contrasting sufficiently good epistemic conditions with ideal epistemic conditions--if so, then his caution that he is not saying that ordinary-language empirical statements can be conclusively verified is compatible with holding that ideal justification is conclusive; (ii) Putnam is identifying sufficiently good epistemic conditions with ideal epistemic conditions--in that case, ideal justification does not amount to having conclusive reasons. I find no firm textual support for either (i) or (ii), and so, with a glance forward, I intend to throw my lot in with (i)--justification under ideal epistemic conditions provides conclusive reasons.

Let's think about Putnam's example. The better the epistemic conditions are in the ways to which Putnam draws attention, the more *unreasonable* it would be to doubt that there is a chair in Putnam's office. If I could doubt that claim under sufficiently good epistemic conditions, what else could I reasonably doubt? A great many things, it seems--especially concerning the reliability of my visually induced beliefs on other occasions. If I can be mistaken, not merely in sub-optimal epistemic conditions, but under especially good ones, then this is a reason to worry about the general reliability of my perceptual belief-forming mechanisms. And if my doubt can reasonably extend to other occasions, then I am calling into question, not just one belief, but a significant portion of my view of the world. "What I hold fast to is not *one* proposition but a nest of propositions."[66] Moreover, many of my beliefs will have inferential links to my perceptual beliefs, and since learning my language has depended in part upon encountering visual samples and paradigms under relatively good epistemic conditions, I may yet have further cause for concern.

Of course, there is no contradiction in my doubting Putnam's claim under "sufficiently good epistemic conditions." Nor is it utterly self-defeating in the way that trying to doubt that I have not always been a brain in a vat is. On the strength of this we can say that it is not really impossible that I am mistaken about there being a chair present in the circumstances Putnam describes. But it is really

---

[65] Being a speaker of English would be an asset to verification in this case.

[66] Wittgenstein (1969), §225.

improbable, and it is unreasonable to harbour such doubts. The conclusion of the argument that we have not always been brains in a vat constitutes, on this picture, an example of an ideally justified contingent statement. We have conclusive reasons for thinking that we have not always been brains in a vat. Doubt here is incoherent (but not *impossible*). In the case of ordinary empirical statements, like the one of Putnam's example, doubt is not incoherent, and epistemic conditions for that statement are not ideal. They are, however, sufficiently good to make doubt unreasonable.[67]

It needs to be reiterated that, even with some of the detail filled in, this account of truth remains a picture, as Putnam desires. My remarks about the "idealization" component of Putnam's formulation should be taken as making perspicuous the intuitions that underlie that formulation. All that I have done is to suggest a way of clarifying what sort of relation holds between justification and truth. To see truth and warrant as internally related is to see truth as a normative, epistemic concept, as a "substantial property" of statements. I add only that to speak of epistemically ideal conditions is to speak of conditions under which it would be incoherent or self-defeating to doubt a particular statement. To speak of good epistemic conditions is to speak of conditions that approximate the ideal, conditions under which it would be unreasonable to doubt a particular statement. None of this *replaces* the obvious claim that "*we have said what is true iff what we were talking about is as we have said it to be*,"[68] but it does remind us that there are no truths apart from the possibility of our giving voice and license to them. The truth of a sentence depends on its meaning and on the way things are--though "things" should not be taken to have any special ontological import. But this should give no comfort to the metaphysical realist, or any other critic of the picture of truth as idealized justification.

---

[67] I draw here on some of Wittgenstein's remarks in *On Certainty*, but I would hesitate to say that the view I have been offering is quite Wittgenstein's. See *ibid.*, *passim*.

[68] Alston (1978-79), p.780. Compare Putnam's italicized remark:

> *[A] statement is true of a situation just in case it would be correct to use the words of which the statement consists in that way in describing the situation.* (Putnam, 1988, p.115.)

Chapter 3: Incommensurability and Conceptual Schemes

I want to distinguish three kinds of relativism: relativism about truth and rationality, respectively, and "conceptual relativism" or relativism about "conceptual schemes." The relativist about truth--and by "relativist" I shall typically mean "cultural relativist"--holds that what is "true" varies from one culture to another. The relativist about rationality holds much the same view about what is rational. In saying these things, such relativists do not just mean that different propositions are *believed* true or rational from one culture to the next, but that the truth-values and rationality of the propositions themselves vary with cultural difference. These forms of relativism often go hand in hand, and I shall treat them as (more or less) one under the heading "epistemic relativism" in Chapter 4.

In this chapter I shall focus on conceptual relativism. This kind of relativism is the most plausible *prima facie*, since it is reasonable to suppose that cultural differences--which are abundant--have something to do with different ways of thinking about the world. If another culture finds plausible and meaningful a range of beliefs and related practices that strike us on first acquaintance as peculiar or even irrational, we might try to explain this difference by suggesting that our two cultures occupy "different worlds," or that our languages divide the world up in different, "incommensurable" ways.

I shall treat Thomas Kuhn's *The Structure of Scientific Revolutions* as representative of conceptual relativism--though, as we shall see, understanding Kuhn's intentions in that book is problematic. My main concern, however, will not be to interpret Kuhn, but to criticize "conceptual relativism." Drawing on the work of Donald Davidson, I will argue that the incommensurability of conceptual schemes is, at best, an idle notion, if we understand it as the presence of a logical barrier to the possibility of understanding one language from the perspective of another. Such relativism, I shall suggest, resembles the classical problem of *other minds*, a problem often associated with (but not only with) metaphysical realism. Vulnerability to Cartesian scepticism seems to imply vulnerability to scepticism about others. Metaphysical realism and conceptual relativism thus appear here as two sides of the same coin, whose value of exchange is limited to haggling in the marketplace of illusion. Given the dichotomous nature of the realism-relativism debate, this should come as no real surprise.

However, Davidson's argument needs revision, for it turns on the claim that we have no conception of truth, apart from our conception of translation. I have claimed that we *do* have another conception of truth, namely, truth as an idealization of rational acceptability. Fortunately, Davidson's insight can be revised to show that we have no idea of rational acceptability independently, not of the notion of translation, but of the more practical concept of *interpretation*. This revision makes sense of the idea that parts of a language might not be translatable, but rules out the idea that a language, or a portion thereof, might be inherently *incomprehensible*. Anything that we can recognize as a language we can *understand* in time, though we may not always be able to translate. We can then admit, where Davidson cannot, the notions of incommensurability and "conceptual schemes," but without the incoherence that accompanies conceptual *relativism*.

I. Conceptual Schemes

Some time ago Agence-France Presse (AFP) reported a number of "mob attacks" in Lagos in which apparently innocent citizens were beaten, sometimes to death, for allegedly stealing men's genitalia.[1] These thefts were thought to have been accomplished by means of "bodily contact such as handshakes" or by asking "the time of day or for directions." According to Reuters, which also reported the alleged theft of women's breasts,[2] medical examinations of complainants showed that "organs were in their natural place and functioning,"[3] but this empirical disconfirmation did not deter those who put faith in the rumour. "Many Lagos residents," said AFP, "now go about the streets checking ... their genitals immediately after a handshake or after bodily contact with a stranger."

The cheeky tone of the AFP-report suggests that we are to regard this tale as one of the eccentricity, incomprehensibility and general lack of "scientific" sophistication of Nigerian culture.[4] When we consider that vigilantism--even lynching--is a "frequent popular reaction to the police's corruption and perceived indifference at the city's high rate of robberies and muggings,"[5] the *extremity* of the reaction to the "bizarre rumour" is less baffling. And if we think of the "superstition" that breasts and genitals can be stolen by casual contact as a metaphor for our own culture's fear of diseases such as AIDS, the whole series of events makes even more sense.

But even if we are critical of the more obvious ethnocentrism of this story, it is difficult to shake the feeling that a belief in the magical theft of genitalia is beyond our abilities to grasp properly. Are the people described simply *irrational*? Do the alleged "thefts" bear any resemblance to what we would normally think of as theft? And, if not, are there any clear--or unclear--criteria for deciding whether or not the thefts really did take place?

The anthropologist, Dan Sperber, tells of a Southern Ethiopian acquaintance named Filate, who "in a state of great excitement" asks Sperber to "kill a dragon" whose "heart is made of gold" and which "has one horn on the nape of its neck."[6] Sperber is bewildered:

> [H]ow could a sound person believe that there are dragons, not 'once upon a time', but there and then, within walking distance? How am I to reconcile my respect for Filate with the knowledge that such a belief is absurd?[7]

---

[1] See "Lagos men," "Nigerians fear" and "Vanishing organs".

[2] See "Sex-organ scare."

[3] See *ibid.*

[4] The Reuters-report is somewhat more reserved.

[5] See "Lagos men."

[6] Sperber (1982), p.149.

[7] *Ibid.*, p.150.

This sort of question might prompt us to embrace what I call in Chapter 5 the "ethical-political argument" for relativism (though Sperber is not so prompted). If we are to make sense of the behaviour of Filate or of the angry crowds in the Nigerian capital, the argument goes, then we must interpret their actions and beliefs as shaped by a very different set of concepts from our own. If we abandon our urge to force our views on them, then we see that their ways of conceiving of the world make sense of these incidents which, to English-speakers with a modern European cultural heritage, for example, seem bizarre. And if we have difficulty grasping how anyone might rationally believe that his penis had been stolen--despite evidence to the contrary--or that nearby there is a golden dragon with a horn on the nape of its neck, then that is just because our "conceptual scheme" has no room for their ways of thinking. In a manner of speaking, we live in "different worlds," and we cannot reasonably fault them for not living in our world. In our world there are no dragons, and the theft of someone's breasts or penis would be a very different and, we are inclined to think, more horrific happening. But Filate inhabits a world in which there *are* dragons, and the angry Lagonians have something real to fear. What are we to make of this position?

In a classic paper of Anglo-American philosophy Donald Davidson tries to undermine the "third dogma" (ITI 189) of empiricism--the dogma that there is a division to be made between a conceptual scheme and its empirical content. In Davidson's view, the only intelligible account of what a conceptual scheme could be is captured by saying that a conceptual scheme is a language (or a group of intertranslatable languages), and the only sense to be made of the idea of *different*, incommensurable conceptual schemes amounts to saying that there could be languages that could not be translated. I shall call this view "conceptual relativism."[8] Such failures of translatability might be complete or partial. But, argues Davidson, a language that we could not possibly translate would not be recognizable *as a language*. So, it is idle to suppose that there could be different conceptual schemes and senseless to speak of there being *one* conceptual scheme.

Whether commensurability is best captured by intertranslatability is a point to which I shall return below, but for now let us accept Davidson's assumption and see where it takes us. As Rorty notes,[9] the scheme-content dogma that interests Davidson is not just a dogma of empiricism; it has a clear precursor in the Kantian synthesis of intuitions in accord with the pure concepts of understanding. Indeed, we might see it as a version of a general distinction between form and matter. But Davidson's paper responded to a trend in English-language philosophy, which he perceived as partly a result of Quine's famous rejection of synonymy and reductionism, empiricism's first two dogmas:

> The dualism of the synthetic and the analytic is a dualism of sentences some of which are true (or false) both because of what they mean and because of their empirical content, while others are true (or false) by virtue of meaning alone, having

---

[8] The choice of terminology is unavoidably arbitrary, since "relativism," "conceptual relativism" and the like receive no consistent treatment throughout the literature on this topic.

[9] PMN 261.

> no empirical content. If we give up the dualism, we abandon the conception of meaning that goes with it, but we do not have to abandon the idea of empirical content: we can hold, if we want, that *all* sentences have empirical content. (ITI 189)

Quine's general strategy is well known.[10] Analytic truths and falsehoods are supposed to comprise a proper subset of the set of logical truths and falsehoods, and they are supposed to do this in virtue of the synonymy of certain of their subject- and predicate-terms--as in the dog-eared example, "A bachelor is an unmarried man." If sense is to be made of analyticity, then sense must be made of synonymy. However, Quine argues, no clear account of synonymy--including such candidates as synonymy "by definition," substitutivity *salva veritate*, and synonymy defined by semantical rules--can be given. Hence, synonymy--sameness of *meaning*--goes by the wayside;[11] but so, then, must the verification-theory of meaning assumed by the logical empiricists. Without bridge-laws that let us reduce theory-sentences to sets of observation-sentences with which they are synonymous the central project of logical positivism collapses. With it also collapses the idea of a theory-neutral language--a language of sense-data or of physicalism that could bridge the gaps between rival sets of theory-sentences, providing a ground for common understanding. As Davidson says, "To give up the analytic-synthetic distinction as basic to the understanding of language is to give up the idea that we can clearly distinguish between theory and language" (ITI 187).

However, being awakened from this dogmatic empiricist slumber does not entail being dragged from the comfy bed of empirical content, and it is this fact that gives the idea of a conceptual scheme a dream-like quality that did not previously attach to it.

The notion of a conceptual scheme had already been entertained by the logical empiricists. In "Empiricism, Semantics and Ontology" Rudolf Carnap suggests that there may be a variety of "linguistic *framework[s]*"--*i.e.*, "systems" of "ways of speaking" that recognize rules appropriate for discussing a "kind of entities."[12] Examples include "thing language" in which we discuss "the spatio-temporally ordered system of observable things and events,"[13] and the frameworks in which we discuss natural numbers, propositions or spatio-temporal coordinates. Within each of these frameworks it makes sense to raise certain questions: "Is there a white piece of paper on my desk?" or "Did King Arthur actually live?" for example. These are *internal* questions--in this case, internal to the linguistic framework of thing-language. Such internal questions, says Carnap, should not be confused with *external* questions, or rather, with external *pseudo-*

---

[10] See Quine (1980), pp.20-46.

[11] The less extravagant idea of *contextual* synonymy, I believe, remains untouched. Synonymy, as Quine considers it, is taken to obtain independently of any particular occasion of use.

[12] Carnap (1956), p.206. See D'Amico (1989), p.32f. for his account of Carnap's influence.

[13] Carnap (1956), p.206f.

questions that purport to address matters of *ontology* before accepting a particular linguistic framework. We might be tempted to ask whether there are any spatio-temporal objects in general. But such a question, says Carnap, is senseless, because its intelligibility presupposes the very linguistic framework that it calls into question. The only way to make sense of the query, he claims, is to regard it as a *pragmatic* question about whether or not we ought to talk in thing-language.[14] We might choose to abandon thing-language in favour of some other "form of language."[15] But to do so is not to exchange one set of ontological commitments for another; it is merely to settle on a new way of talking, and this is something that we do for pragmatic reasons, such as "efficiency, fruitfulness and simplicity."[16]

Unlike the conceptual schemes of the conceptual relativist, Carnap's frameworks do not put up barriers to understanding, because regardless of which frameworks we use, we can always ground communication across frameworks on what is objectively given us in experience. Observations may not uniquely confirm a specific linguistic framework, but they *are* given, regardless of one's choice of frameworks. And although Carnap allowed for the possibility of different linguistic frameworks even at the level of observation-statements, he also endorsed the possibility of determinate translation from one observation-language to another.[17]

However, this idea of a theory-neutral set of intertranslatable observation-languages and the sharp distinction between "questions of meaning and questions of fact,"[18] as Carnap acknowledged, rest on the notion of a strict synonymy of terms in different observation-languages[19] and "an absolute distinction between the analytic and the

---

[14] See *ibid.*, p.207f.

[15] *Ibid.*, p.208.

[16] *Ibid.*

[17] Carnap came to prefer a language of physics to one of sense-data for observation-statements, but not because he saw the choice as a "cognitive" one. Physicalist language is preferable to phenomenalist language because of its "intersubjectivity, *i.e.*, the fact that the events described in this language are in principle observable by all users of the language" (Carnap, 1963, p.52). But this advantage is of a kind with the criteria of "efficiency, fruitfulness and simplicity." Physicalism was for Carnap a question of "'attitude' and not 'belief' because it was a question of practical preference, not a theoretical question of truth" (*ibid.*, p.51).

[18] Carnap (1956), p.215n5.

[19] Responding to Quine's and White's (see White, 1950) criticisms of the analytic-synthetic distinction in "Meaning and Synonymy in Natural Languages," he observes that an Anglophone linguist who tries to understand a German predicate through radical translation must assign a property to that predicate as its intension. This assignment

> ... may be made explicit by an entry in the German-English dictionary, conjoining the German predicate with an English phrase. The linguist declares hereby the German predicate to

synthetic,"[20] between "logical and factual truth."[21] Abandoning these two dogmas of empiricism requires that we forego theory-neutrality, and regard scientific hypotheses as having ontological import.

But even if we reject the dogmas, we may be tempted to retain our old talk of frameworks or of "the conceptual scheme of science"[22] without the benefit of a theory-neutral observational bridge between alternative frameworks. Thus, Quine says that his "pragmatism" is "more thorough" than Carnap's which "leaves off at the imagined boundary between the analytic and the synthetic."[23] It is not just scientific hypotheses that are implicated in choosing a "convenient conceptual scheme or framework for science,"[24] but our whole ontology. Faced with the changing flux of empirical content, bereft of a neutral language of sense-data or physics, we need our schemes or frameworks to "*organize*" or "*systematize, divide up* (the stream of experience) ..." (ITI 191). And there are diverse ways to do this organizing.

If "Two Dogmas ..." suggested an important revision of the idea of a conceptual scheme, its dry (though lucid) discussions of definition, interchangeability and semantical rules still required a little help to become influential in Anglo-American philosophical circles. It is fair to say, I think, that help came from the engaging historical studies of such philosophers and historians of science as Thomas Kuhn.[25] Kuhn, who cites Quine as an important influence,[26] is known for disputing a traditional conception of scientific-theory change, according to which science is cumulative or accretive--*i.e.*, according to which scientific discoveries are really *discoveries*, and when old theories are rejected in favour of new ones, those new ones can tell us new things about the objects studied by the old theory.

The apparent continuity in the history of science, is, Kuhn argues, largely an illusion brought about by scientists and historians of science, who tend to view the science of their day as constituting *progress* over the science of earlier times. There is great resistance to theoretical change in the natural sciences, and to understand this resistance we cannot treat the history of science as one of continuous dialogue against a background of shared criteria of relevance. So-called "crucial experiments," Kuhn points out, are seldom performed to settle a current dispute between two theories that aim to explain

---

[19] be synonymous with the English phrase. (Carnap, 1956, p.237)

[20] Quine (1980), p.45f.

[21] Carnap (1956), p.215n5.

[22] Quine (1980), p.44.

[23] *Ibid.*, p.46.

[24] *Ibid.*, p.45.

[25] But the theme is widespread. The tendency of philosophers of many different stripes, and of linguists, to see language as a formal system, quite apart from its "pragmatic" aspects, encourages, I think, theses of incommensurability in the sense that Davidson intends.

[26] Kuhn (1969), p.vi.

roughly the same phenomena. Such experiments are usually carried out after a new theory has already been adopted, and it often remains open for the holder of an obsolete theory to interpret the results as confirming, rather than falsifying the old approach, assuming that advocates of the older "paradigm"[27] have not all died off.

As such, theorists in opposing camps often find themselves talking at cross-purposes--if they bother to talk with each other at all--for, it seems, they cannot even agree upon their standards of warranted assertibility. This communicative gulf is broadened by the fact that terms from an older theory can survive to play a role in the newer theory, but often the role will be slightly different, suggesting that, for example, what Thomson meant by "atom" is not what Bohr meant by "atom." In short, it may seem that practitioners of one paradigm are not even talking about the same things as are practitioners of another paradigm. This casts into doubt the idea that scientists gradually acquire more and more knowledge about the same things with each theory-change. A cumulative account of science applies only to "normal science," research carried out *within* a paradigm.

Now, the communicative breakdowns that separate scientists in different traditions or eras Kuhn imputes to the incommensurability of their respective paradigms. When he says this he seems at times to have in mind a failure of translatability.[28] This raises the question --dwelt on by Davidson--of whether the failure in question is a partial one or a complete one. Some passages from *The Structure of Scientific Revolutions* suggest that it is the latter kind:

> In so far as their only recourse to [the] world is through what they see and do, we may want to say that after a revolution scientists are responding to a different world.[29]

If scientists following a new paradigm live in "a different world," then similarities of the vocabulary that they retain are no more than that--the reference of the terms involved will have shifted with their new situation in the context of the new paradigm. The atoms that Thomson investigated were not the atoms that Niels Bohr investigated--indeed, they were not even components of the same world. This account of Kuhn's position commits him to conceptual relativism, for if the referent of a term changes, along with its meaning, we are left with no reason to regard the concept involved as a concept of the same thing--no reason to think of it as the *same concept*.[30]

---

[27] Kuhn's use of the word 'paradigm' is, of course, infamously ambiguous, but I do not intend to explore that issue of interpretation here. I shall use the word in fairly innocuous contexts.

[28] This initial estimate will be revised below.

[29] Kuhn (1969), p.111.

[30] Implicit here is a "criteriological" account of the reference of concepts. According to such an account, the reference of a concept is given by the meaning of the concept, and the meaning of the concept is similar to a disguised description--much like Russell's account of proper names as disguised descriptions. On Russell's reckoning, we are to understand the meaning of, *e.g.*, the name 'Walter Scott' to be "the

But now, if concepts, terms, reference and meaning change utterly from one paradigm to the next, then what we knew--what was *true*--before the shift of paradigms, seems no longer to be true. Either those former truths are now false, or they are neither true nor false, because they purport to talk about entities peculiar to that erstwhile world of the pre-revolutionary scientist. Truth--it seems--is relative to incommensurable conceptual schemes, and anything that we are epistemically justified in saying or believing is similarly relative.

II. Davidson's Argument

The position just outlined is one for which Davidson does not have much sympathy, and his reasons for rejecting it are *prima facie* quite compelling. Davidson uses two general strategies to argue against the possibility of complete failure of translatability. One is to cut away at the metaphors used to illustrate what it would mean for there to be different conceptual schemes. The other strategy is to consider what would constitute our having a good reason to suppose that we had encountered speakers of a language that was completely untranslatable. Davidson suggests that nothing could count as such a reason, because any evidence that would count in favour of such a hypothesis would also count in favour of believing that the behaviour manifested by these alleged speakers "was not speech behaviour." "If this were right," he continues, "we probably ought to hold that a form of activity that cannot be interpreted as language in our language is not speech behaviour" (ITI 185f). We would be in no position even to recognize an untranslatable language as being a language.

The possibility that a language might be, in principle, beyond our powers of translation thus appears similar to the Cartesian scepticism that we encountered while examining metaphysical realism in Chapter 2. Indeed, it is tempting to say that the conceptual relativist is a kind

---

[30] author of *Waverly*," or perhaps a conjunction of descriptions. (See Russell, 1919.) Similarly, a criteriological account of the reference of, *e.g.*, the concept "gold" would see the reference of 'gold' as fixed by a description with which 'gold' is taken to be synonymous. So, to oversimplify, we could suppose that in Kant's day 'gold' meant "a yellow metal," but that now 'gold' means, perhaps, "the element with atomic number 79." Plainly, these two descriptions do not pick out precisely the same things, though there may be some overlap.

I think that not just *any* change in use constitutes a change of *literal meaning*. Every occasion on which a word is used is different in some way or another from every other occasion, but ordinarily we do not think of this as constituting a difference of "meaning," since we take "meaning" here to be "literal meaning," or (roughly) the broadest distinguishable type-description of the use to which the word is usually put. Generally, there will be no *rule* for deciding when the "same concept" is in use in two different "schemes." Each case will require examination and maybe argument about whether differences of use seem great enough to merit a judgment of "different concepts." What is crucial for understanding is that we can trace the historical changes of use and interpret the earlier use in light of what we can learn about those historical changes.

of *realist*--a realist about conceptual schemes.[31] Giving in to this temptation reveals the problem for the conceptual relativist: there is no way to get outside of a conceptual scheme while one claims that there are such schemes (in Davidson's sense). Any such claim is itself "trapped inside" a conceptual scheme, like the Cartesian sceptic locked inside the theatre of ideas. Just as metaphysical realism requires that we might be completely mistaken about the nature and existence of the external world, the hypothesis of incommensurability as untranslatability entails that there might be "worlds of meaning" from which we could be logically excluded. An even more appropriate comparison is with the problem of other minds.[32] From the Cartesian sceptic's point of view, the illusion of other minds is, it seems, just as good an explanation as the actual existence of other minds. Any reason to doubt the existence of an external world seems a good reason to doubt the existence of other minds, too.[33]

As we saw in the preceding chapter, however, the idea that the Cartesian sceptic might be right, that the external world might simply be an illusion, is self-defeating. The best account that we have of the sceptic's doubt is one that casts doubt on the intelligibility of that original doubt. If the external world is an illusion, then our words of doubt do not mean what we think they mean; hence, it could not really be the case that the external world is an illusion. But the same account of meaning and reference undermines sceptical doubt about other minds and the conception of objectivity that enables the illusion of that sceptical doubt. That there are other minds--or, better, other language-users--is no more and no less certain than that I am a language-user. The behaviour that they manifest is the same sort of behaviour that I manifest. Any reason that I could have for thinking that such behaviour in another were not the behaviour of a language-user would also be a reason for thinking that I myself am not a language-user. But this I cannot intelligibly doubt.

This is not an argument from analogy. I am not saying that from a Cartesian starting-point I can judge that there are other minds, because the behaviour and appearance of certain other bodies is enough like my own to warrant an inference. What I am saying is that I can be mistaken about my own mind, just as I can about those of others. At the same time, there is nothing that in principle is hidden from me about the minds of others or about my own mind, though I can make mistakes about either and though others may well choose to conceal their intentions on occasion. (I may also deceive *myself*.)

In dispensing with metaphysical realism, as I have, I dispense also with metaphysical realism about other minds, a doctrine that implies

---

[31] Putnam comes close to saying this at one point. See PP3 238. I shall take up this point at greater length in Chapter 4.

[32] See RTH 124 and PP3 236ff. See also Root (1986).

[33] Descartes claims explicitly that knowledge of other people, like knowledge of the external world, is *inferential*:

> [I]f I look out the window and see men crossing the square, ... I normally say that I see the men themselves .... Yet do I see any more than hats and coats which could conceal automatons? I *judge* that they are men. (CSM II 21)

that we might really know absolutely nothing about other minds, that the existence and nature of other minds are independent of our epistemic capacities. Only if we have to entertain this sceptical threat, does it make much sense to worry further that there might not *be* any other minds.[34] As Sartre writes:

> The Other's existence will always be subject to doubt, at least if one doubts the Other only in words and abstractly, in the same way that without really being able to conceive of it, I can write, "I doubt my own existence."[35]

Just as it is logically possible that I should not exist, or that we should all be brains in a vat, it is logically possible that I should be the only mind in the universe. But a logical possibility is no reason for genuine doubt; only a *real* possibility can be that. The philosopher "doesn't doubt when he says he's not sure";[36] the philosopher doubts "only in words and abstractly." Any reason that I could have for doubting the existence of others is a reason for doubting that I understand the words with which I formulate my solipsistic fear and a reason for doubting my own status as a mind or a language-user. For the words that I use acquire their meanings from a complex set of practices that I learn *from others*. Or--to put the point with less prejudice--words have meanings because they are *used*, and to suppose that apparent others might behave exactly as they do without its being the case that they *really* use words is an absurd hypothesis. It is to suppose that there is some mysterious criterion that *I* satisfy, but which other apparent language-users do not, "that there is a fundamental difference between my mind and others' minds, that there is a distance between us to be overcome."[37]

The force of these considerations will become clearer below, for Davidson's line of argument against conceptual relativism is similar. Any reason for thinking that purported linguistic behaviour is in principle unintelligible to us is a reason for supposing that it is not linguistic behaviour. Consider how this reasoning is supposed to work. Conceptual relativism entails epistemic relativism--though this is not at first apparent. What is supposed to be different about

---

[34] This is not to say that just any rejection of metaphysical realism relieves us of the threatened burden of solipsism. Kant's arguments--with the possible exception of the Refutation of Idealism--leave the existence of other minds, not to mention their possession of a uniform set of pure concepts of understanding, purely a matter of faith. In a sense, Kant remains a metaphysical realist about other minds, as Sartre observes: "It is in fact by this position [Kantianism] with regard to the existence of the Other that we suddenly explode the structure of idealism and fall back into a metaphysical realism" (Sartre, 1956, p.311).

[35] *Ibid.*, p.337.

[36] Wisdom (1957), p.174.

[37] Overall (1988), p.96. This is not to deny the phenomenon of first-person authority. See, *e.g.*, Davidson (1984b). I shall offer a qualified endorsement of Davidson's position in Chapters 7 and 8.

holders of another conceptual scheme is not that they are wrong about the world, whereas we are right, but that many of their views are as true as our own *and* incommensurable with our own. So, an alternative conceptual scheme is one containing a significant number of true sentences, which are not amenable to translation. But does this make sense? Do we "understand the notion of truth, as applied to language, independent of the notion of translation" (ITI 194)?

Davidson contends that we do not--that the only conception we have of what *truth* is is embedded in something like a Tarski truth-definition for a language. We give such a definition for our own language by the simple process of disquotation for each sentence of our language, as in 'Snow is white' is true, if and only if snow is white--to use another well-worn example. But while this seems a trivial exercise for our own language,[38] when we turn to another language we are immediately forced to become translators. To give a truth-definition for L, I must translate each sentence of L, giving its truth-conditions in my own language, and no sense can be made of the notion of truth beyond this exercise. Hence, supposing that a language could be utterly resistant to the possibility of translation --that there could be a conceptual scheme incommensurable in this sense with our own--is futile. According to Davidson, nothing could be recognized as a language if I could not take its components to be bearers of truth-values, and I can do that only if I can translate it.

At this point a reply is likely to occur to the conceptual relativist. Why not try to accommodate Davidson's criticism by retreating to a position that insists upon only a partial failure of translatability? Of course, she might say, there must be an assumption that some sentences of a language are translatable, if we are to recognize it as a language, but that still leaves open the possibility of a sphere of discourse that resists any effort we could ever make to understand it. Serious talk of dragons and of the theft of genitalia comprise such spheres perhaps.

Indeed, Kuhn might be thought to hold just such a view, sometimes recognizing that his claims about living in "different worlds" need to be put more weakly.[39] He also tells us that a more traditional picture of what happens when scientists embrace a new theory "can be neither all wrong nor a complete mistake" and locates the difficulty of making his point in the failure of contemporary epistemology to provide a vocabulary adequate to his claims.[40]

---

[38] Or so I shall regard it here. I have strong reservations about the value of such a procedure (assuming it is even possible) for the task of giving an account of linguistic meaning.

[39] I am ... aware of the difficulties created by saying that when Aristotle and Galileo looked at swinging stones, the first saw constrained fall, the second a pendulum. The same difficulties are presented in an even more fundamental form by the opening sentences of this section: though the world does not change with a change of paradigm, the scientist afterwards works in a different world. Nevertheless, I am convinced that we must learn to make sense of statements that at least resemble these. (Kuhn, 1969, p.121)

[40] See *ibid.*

However, the concession that failure of translatability may be only partial does not save the idea of a conceptual scheme from problems, according to Davidson. His argument has the form of a dilemma. Either the incommensurability of conceptual schemes amounts to a complete failure of translatability--in which case the alleged holders of an alternative scheme cannot be thought of as speakers at all--or the incommensurability in question amounts to only a partial failure of translatability--in which case, there is only a *difficulty of translation*, not a failure of translatability. Let us test the point of the second horn of this dilemma.

Davidson's reasoning goes like this. Anything that would count as evidence for a partial failure of translatability would also count as evidence for mere disagreement or commonplace misunderstanding between field-linguist and speaker. According to the principle of charity, translation cannot even get off the ground, unless we assume that we share a great many beliefs with the speaker whose utterances we are trying to translate. We can attribute no meanings without assuming beliefs, and we will make no progress at all, if we begin by ascribing systematically beliefs that we reject. "Since knowledge of beliefs comes only with the ability to interpret words, the only possibility at the start is to assume general agreement on beliefs" (ITI 196). And since we typically take ourselves to have true beliefs, we must suppose that many of our interlocutor's beliefs are also true.

> The guiding policy is to do this as far as possible, subject to considerations of simplicity, hunches about the effects of social conditioning, and of course our common-sense or scientific knowledge of explicable error. (ITI 196)

We ascribe certain meanings to the speaker in light of the assumptions we make about her beliefs and then attribute further beliefs and meanings, revising each in the light of the other. The process is one of arriving at a coherent balance of these factors. If I ascribe certain beliefs to a speaker, I will tend to construe her utterances in some ways more readily than in others. And the interpretations that I choose will, in turn, constrain what other beliefs I ascribe to her.

Indeed, it is a mundane procedure that we resort to within our own languages, whenever there is some question as to the meaning of someone's utterance. If a French-speaker says, "Shut the door!" the beliefs that I attribute to her about my fluency will affect whether I interpret her exclamation as a command or as a declaration of love. Confronting the speaker of a completely unfamiliar language is simply a more sophisticated case of the same kind.

But having been compelled by charity to grant considerable agreement of belief, it becomes difficult to imagine what could force on us the conclusion that beyond a core of shared belief there lies a murky region of conceptual relativism in which understanding breaks down:

> If we choose to translate some alien sentence rejected by its speakers by a sentence to which we are strongly attached on a community basis, we may be tempted to call this a difference in schemes; if we decide to accommodate the evidence in other ways, it may be more natural to speak of a difference of opinion. (ITI 197)

The alleged incommensurability of schemes reduces to a difference of belief. That it should be *tempting* to view the difference between, say, the medievals and Galileo as one of conceptual schemes comes as no surprise on this account, for the discovery of the moons of Jupiter and the phases of Venus implied *either* the falsehood of an embedded set of propositions, relating humanity, the earth, and the universe to God--the creator of a *harmonious* world--*or* a clever and elusive new interpretation of Galileo's claims, which would see them as compatible with the alleged harmony of nature. Since the latter was not quickly forthcoming, and the former was simply too upsetting, Galileo had to be wrong *a priori*. That, however, is just to say that Galileo and the Church had a very basic difference of opinions, not of conceptual schemes. Our rejections of Filate's belief in dragons and the rioters' fears about genital-theft are to be similarly understood.[41]

Davidson's general point comes into focus when we realize that incommensurability as untranslatability, like Quine's indeterminacy of translation, "begins at home"[42]--or *would*, if it could begin anywhere. Let us put ourselves in the position of the speaker who, supposedly, has access to a sphere of discourse whose concepts we cannot make out

---

[41] D'Amico dismisses Davidson's argument, claiming that Davidson implicitly "leans on the myth of the given" (D'Amico, 1989, p.130). His assertion rests on the supposition that Davidson is arguing for some kind of "knowledge by direct acquaintance" (*ibid.*), a suspicion that he supports by appealing to Davidson's final remarks:

> In giving up the dualism of scheme and world, we do not give up the world, but re-establish unmediated touch with the familiar objects whose antics make our sentences and opinions true or false. (ITI 198)

However, it is curious to treat this passage as suggesting that Davidson is committed to some kind of unproblematic, immediate, incorrigible relation of ourselves to the data of our senses. That is to ignore remarks such as the following:

> Nothing ... no *thing*, makes sentences and theories true: not experience, not surface irritations, not the world, can make a sentence true. *That* experience takes a certain course, that our skin is warmed or punctured, that the universe is finite, these facts, if we like to talk that way, make sentences and theories true. But this point is put better without mention of facts. The sentence 'My skin is warm' is true if and only if my skin is warm. Here there is no reference to a fact, a world, an experience, or a piece of evidence. (ITI 194)

It is integral to Davidson's position that true sentences and true beliefs do not get that way in virtue of their reference to a reality that can be given a theory-neutral description. In "Reality without Reference" (ITI 215-225) he writes, "Reference ... drops out. It plays no essential role in explaining the relation between language and reality" (ITI 225). However, it is precisely such reference that is presupposed by any position which "leans on the myth of the given."

[42] Quine (1969), p.46.

at all. If we share with her a core of understanding, but cannot in principle understand some set of incommensurable concepts, why should we suppose that she could? If I share a core of understanding with you, and if you with your conceptual resources cannot grasp some set of incommensurable concepts to which I claim access, why should we suppose that *I* could so grasp them?

A language that could in principle be understood only by *some* people, members of one *particular* culture, is not that different from one "which only I myself can understand" (PI §256).[43] Meaning, I suggested in Chapters 1 and 2, is a normative notion: there are correct and incorrect ways of using particular words and phrases, and to understand a word or phrase at all I must be able to recognize, at least sometimes, when it is used correctly and when incorrectly. I understand an expression when I recognize what does and what does not conform with the practice or institution of its use. But if members of some culture can come to understand such a practice, how is it that I could not? How is it that *they* could not come to understand the practices that determine the meanings of *my* words? Perhaps my attempts to involve myself in their practices would be inept or lacking polish, just as my attempts to play hockey or billiards might be inept. However, it would be absurd to suppose that some logical barrier made me a poor hockey-player or a failed pool-shark.

We could, perhaps, imagine beings whose sensory capacities were so unlike our own that we could not recognize them as responding with language to a phenomenon that we could not ordinarily observe. But only if their entire language were so affected--only if no phenomena accessible to them were accessible to us--would their language be completely incomprehensible from the perspective of our language, and it is difficult to imagine what this latter possibility would be like.

Less extravagantly, there are important differences of experience produced by, *e.g.*, anatomical differences. Women and men typically stand in different relations to sexuality and reproduction (and to be sure, there are many differences *among women* and *among men*), and most cultures lay atop these differences many other social constraints on our experiences. To take a less important example: being partially colour-blind, I can see things that some others cannot, and cannot see things that some others can. If someone has had experience vastly different from our own, then it may be that no amount of translation can get at those privileged experiential data--at least not clearly. But that does not rule out the possibility of our coming to understand such a person, if and once we ourselves have similar experiences. The present point is one of how our speaker could understand herself as a speaker in, or user of, that other conceptual scheme, from the vantage of that core of understanding that we share. To insist on partial incommensurability is to embrace a kind of scepticism, not just about other minds, but about *one's own mind*. It is to suppose that there are facts about oneself that are, not just in practice, but logically

---

[43] As Derrida puts it, "A writing which is not structurally readable--iterable--beyond the death of the addressee would not be writing" (Derrida, 1988, p.7).

beyond one's understanding.[44] Moreover, it is to display an even stronger kind of epistemic neurosis than we met with in the preceding chapter, for at the same time as this doubt is incurred by realist and conceptual relativist alike, each must ignore it in order to proceed.

III. Truth, Interpretability and the Principle of Humanity

Davidson's argument against the very idea of a conceptual scheme has generated much discussion, but a survey of some of the many responses to it reveals a curious phenomenon. While one set of opinions tends to converge on the claim that the argument as it stands will not do, there is disagreement over whether this signals the failure of his position, or simply the need for revision. The views that emerge from these contrasting assessments seem to have a great deal in common. Alasdair MacIntyre, for example, thinks that there is something fundamentally wrong-headed about Davidson's approach here. He fears that it thrusts on us a kind of illusion:

> [A]ntirelativism pictures us first as necessarily inhabiting our own conceptual scheme, our own *weltanschauung* ... and second as necessarily acquiring whatever understanding we may possess of the conceptual schemes and *weltanschauungen* of others by a process of translation so conceived that any intelligible rendering of the concepts and beliefs of the others must represent them as in all central respects similar to our own.[45]

MacIntyre's fear, it seems, is that this illusion translates into yet another wielding of cultural hegemony, a wilful ignorance of cultural differences elevated to the level of a methodological procedure and a logical principle. Similarly, Lorraine Code concedes that it may be possible as a matter of logic "for all linguistic utterances to be translated into any other language without remainder," but she is

---

[44] My use of "understanding" here begs certain questions about the interpretation of Davidson and, possibly, about the relation between translation and understanding. I shall take up these issues, especially the latter, in Section III. But it needs to be said that the conceptual relativist usually does not mean to say that it is logically or really impossible *simpliciter* to understand the concepts of a scheme incommensurable with one's own. The point is just that given one's commitment to conceptual scheme A, one cannot understand conceptual scheme B using A-concepts. It might be really, or at least logically, possible, on this view, for an A-schemer to understand B-concepts by immersing herself in B-culture and *ignoring* A-concepts. But any attempt to draw on A-concepts to understand B-concepts, *e.g.*, by comparison and analogy, is doomed to failure, and insofar as an A-schemer cannot set aside her A-concept bias, she will not understand B-concepts. Implicit here are accounts of concepts and of standards of rationality as having applications that are strictly delineated in terms of necessary and sufficient conditions. I criticize these views later in this chapter and in Chapter 4.

[45] MacIntyre (1987), p.404.

anxious that we not forget that "in practical terms, conceptual barriers are often difficult to assail."[46]

Yet, the views of both these thinkers are similar to the views that I have been advocating here, views that in my version owe much to the seemingly unlikely duo of Wittgenstein and Putnam. Indeed, Putnam argues for a sympathetic revision of Davidson (RTH 113-119), replacing Davidson's "principle of charity" with the principle of "humanity" (RTH 117).[47] Hacking proposes a similar modification, remarking that "since Davidson's argument may seem founded upon a lack of concern for alternative interests, we may fear his premises while we accept his conclusions."[48] And Rorty regards it as "unfortunate" that Davidson "misinterprets Kuhn as meaning 'untranslatable' by 'incommensurable'" (PMN 302n35), but endorses the argument anyway.[49]

What criticism do these philosophers share, but view so differently? To see this it is useful to return to Davidson's confrontation with the conceptual relativist. Davidson's argument relies on the claim that we have no idea of truth independently of our idea of translating another language into our own. Why can't the relativist respond that we *do* have an idea of truth independently of a Tarski truth-definition and that, therefore, sense can be made of the idea of incommensurable conceptual schemes and the accompanying relativism about truth and rationality? The relativist might claim that we have an idea of truth, apart from the notion of translation, as the "warranted assertibility" or "rational acceptability" of sentences. Such a relativist might accept an identification of truth with warranted assertibility or a reduction of truth to an *idealization* of warranted assertibility. In the latter case her position would differ from Putnam's in its reductionism, and it would assume that we can make sense of this idealization of warrant without being able to translate the language to which it belongs. The warrant in question might be construed as springing from, *e.g.*, community acceptance. I am warranted in asserting that snow is white, if my fellows respond favourably to such a claim. Does this kind of strategy avoid Davidson's critique?[50]

To treat truth as connected with warranted assertibility certainly *seems* to be to offer an account of truth different from Davidson's,[51] but if this move is to be successful, then we must also have some notion of warranted assertibility that is independent of our notion of translation. A plausible case seems available for this claim. The notions of warrant, justification, rational acceptability and the like have a crucial role to play in our ability to *understand* declarative sentences. To use a declarative sentence in the context of "fact-

---

[46] Code (1991), p.59.

[47] See also Lukes (1982), pp.262ff. and Grandy (1973).

[48] Hacking (1982), p.62.

[49] PMN 306-311.

[50] Such a move is tantamount to viewing conceptual relativism as dependent upon relativism about truth and rationality, in contrast to the strategy considered above of making conceptual relativism prior.

[51] Rorty disagrees. See PMN 308, and Rorty (1991a), pp.126-150.

stating" discourse often just is to make a claim that can be contested or supported. Declarative sentences have other uses as well--"Yon Cassius has a lean and hungry look" spoken on stage--but anyone who did not understand the connection between warrant and the use of a declarative in fact-stating discourse would not understand that use of a declarative. Indeed, I think it would be mysterious to such a person why "declarative" is used to describe such sentences--and this would be the *least* of her worries. Learning such normative notions as warrant and rational acceptability is intimately tied to the learning of language. This is hardly surprising, since linguistic meaning is normative: there are right ways and wrong ways to use words and phrases, and one's ability to *speak*--or to write or to sign--depends on one's ability to recognize some uses as right and others as wrong.

So, if the idea of warranted assertibility depends upon the idea of *translation*, then our ability to learn *language*--not just *a* language--depends upon our having some notion of translation. But, surely, the consequent here is false. Even if learning another language often involves translation, I do not *have* to translate a language to learn it, since, as Howard Sankey observes, "it is possible to learn a new language by the method of direct immersion."[52] Moreover, I do not learn my first language by translating it. "First language acquisition is not a translation process, and nothing that is absent here can be a necessary ingredient in subsequent learning" (RRSK 37). I begin with the training of my caregivers, and as I learn my language, I learn rough standards of warranted assertibility attached to certain kinds of statements by the community(ies) of speakers to which I belong.

In fact, it is unclear just what to make of Davidson's position at this point. The scenario of translation to which Davidson appeals in arguing against conceptual relativism is a recurrent theme of his work, but it is another question whether his work of this period really treats translation as necessary for linguistic understanding. In a paper that appeared not long before "On the Very Idea of a Conceptual Scheme" Davidson suggests that being able to translate a language (by means of what he earlier called a "theory of meaning")[53] is not *necessary* for linguistic understanding, but *sufficient*:

> Having identified [a speaker's] utterance as intentional and linguistic, we are able to go on to interpret his words ... . [What] could we know that would enable us to do this? How could we come to know it? The first of these questions is not the same as the question what we *do* know that enables us to interpret the words of others. For there may easily be

---

[52] Sankey (1990), p.7. According to Sankey, Davidson confuses *translation* with *interpretation*: the argument against conceptual schemes treats *understanding* or *interpreting* a language, as if it were nothing but translating--or more exactly, as if translation were a necessary condition of understanding and interpretation. (For a similar criticism of Lukes (1982) see RRSK 36f.) Sankey's criticism is, in respects, like Dummett's charge that Davidson's position provides only a "modest" theory of meaning, not one that also accounts for linguistic understanding. Dummett withdraws the criticism later in the same essay. See Dummett (1975).

[53] See ITI 304-323.

> something we could know and don't, knowledge of which would suffice for interpretation, while on the other hand it is not altogether obvious that there is anything we actually know which plays an essential role in interpretation. (ITI 125)[54]

The significance of this passage, in part, is to free Davidson of any commitment to the idea that a speaker must have "implicit knowledge"[55] of a theory of meaning for a language, in order to be able to speak that language. Such a theory of meaning would be roughly, a Tarski truth-definition for the language, which we saw implicated above in Davidson's initial objections to conceptual relativism. But this poses a problem: if it is "not altogether obvious that there is anything we actually know which plays an essential role in interpretation," then it would be odd for Davidson to claim that the notion of translation is a necessary condition for being able to interpret a language. That Davidson calls his paper "Radical Interpretation," not "Radical Translation," further complicates matters. And as other passages show, Davidson moves freely back and forth between "translation" and "interpretation": "...we cannot take even a first step towards *interpretation* without knowing or assuming a great deal about the speaker's beliefs" (ITI 196; my emphasis). Moreover, his conception of truth, as Rorty emphasizes,[56] seems close to warranted assertibility in some places: "We get a first approximation to a finished theory by assigning to sentences of a speaker conditions of truth that actually obtain (*in our own opinion*) just when the speaker holds those sentences true" (ITI 196; my emphasis).

There is, I believe, a way out of the dense undergrowth that is entangling us, for when we re-trace our steps and consider Davidson's argument from another direction it takes on a new appearance. Let us grant the relativist that we have an idea of warranted assertibility independently of our idea of translation--this seems quite correct to me. But it should offer no solace to the relativist, for we do *not* have an idea of warranted assertibility independently of the ideas of *interpretation* and *understanding*. This means that we can reconstruct Davidson's argument in a way that undermines the incommensurability of conceptual schemes in the following way: there is no "conceptual scheme" that we really could not come to *understand*, even if there were no good *translation* of its concepts into our "conceptual scheme."

Let me clarify briefly what separates understanding, interpretation and translation here. In Chapters 1 and 2 I argued that semantic notions like meaning and truth are normative, intentional notions. To the extent that meaning is internally connected with understanding, the latter notion is likewise normative. When I understand a word or a more complex expression, I understand how it is used, and its use is its meaning. There are correct and incorrect ways of using words, and when I understand words, I can tell, more or less, when they are used correctly and when incorrectly. Very often, in everyday conversation,

---

[54] It is tempting to respond as Morton White did to "rational reconstructions": "[H]ow do we establish when people behave *as if* they had done something which they haven't done?" See White (1950), p.322f.

[55] Dummett (1976), p.70.

[56] See PMN 295-311.

for example, such understanding is fairly direct--I hear the words of another, and I understand what she says. Understanding in such cases requires neither translation nor interpretation.[57] However, this sometimes does not happen. With the possibility of understanding goes the possibility of misunderstanding. In such cases, if I am to *achieve* understanding of another, I must resort to interpretation. I must form hypotheses regarding the point of her utterances and then determine whether those hypotheses make sense of her behaviour in the context of whatever else I may know about her that seems relevant. When I do so, I am not necessarily concerned with the Quinean task of constructing a translation-manual for her language, even if we do not share a language.[58] Rather, I am interested in making sense of her utterances on a particular occasion. If we share a language to which the words of her utterances on that occasion belong, then I shall assume our mutual competence with that language as background for my interpretation. If our languages differ, then I can fall back on cultural similarities of which I am aware, or at least on our common humanity--on the fact that we have similar bodies and sensory organs, similar basic needs for food and drink, shelter, some degree of affection, and so on. In both cases, interpretation is local, rather than global, but it can always tend toward the global, as I interpret another on numerous occasions.

Interpretation is not only local, but connected with non-linguistic behaviour and practical affairs. J. E. Malpas links the two points succinctly: "Interpretation is, one might say, always a practical task oriented towards the particular, rather than a theoretical one oriented towards the universal."[59] Davidson sometimes portrays his project as elucidating the idea of radical interpretation, but in borrowing from Quine the scenario of radical translation, he sometimes seems pushed in the direction of describing interpretation as an attempt to provide a "total theor[y]" (ITI 241) for understanding a person.[60] Here a total theory seems analogous to the comprehensive translation-manual that Quine's field-linguist tries to compile. As well, although the notion of radical interpretation is supposed to add a pragmatic constraint (see ITI 125-139) to Davidson's early claims that linguistic competence might be represented by a truth-theory for a language, and although Davidson has written extensively on action-theory (which contributes to the development of "total theories"), interpretation is practical in a way that is not done justice by Davidson's account.[61] In brief, interpretation must be practical, because *meaning* is rooted in practical activity. This was illustrated in Chapter 2, §I, by Putnam's "Turing-test for reference." That thought-experiment showed that for computer-imitations of human language-use to be genuine, meaningful language-use, the machines

---

[57] I shall return to such cases when comparing and contrasting knowledge of the self and knowledge of others in Chapter 7, §§IV-V.

[58] See Quine (1960), pp.27ff.

[59] Malpas (1992), p.112.

[60] This tendency meshes with his interest in "theories of meaning." For a critique of "total theories" see *ibid.*, pp.110-115.

[61] See §V below and Chapter 8, §I for elaborations of this claim.

involved would have to be able to interact with their environment--to engage in practical activity. Our attempts to interpret them would fail, if we could not refer to any such practical activity of theirs in trying to understand them. In order for linguistic behaviour to *be* linguistic behaviour it must be embedded in a background of other behaviour, of practical activity, of ways of life.[62]

The third respect in which interpretation, as I understand it, differs from translation is the respect that I began by mentioning above--its emphasis on warrant, rather than on truth. That Davidson insists that the attribution of truth to the utterances of others is the central requirement of the principle of charity might well be traced again to his debt to Quine. For Quine, as we shall see in the next section, radical translation gets going only because we can form genuine hypotheses regarding which of a speaker's utterances signal the basic attitudes of assent and dissent (WO 29f). For Davidson, who "despair[s] of behaviourism" (ITI 231), these attitudes are transformed into attitudes of "holding true" (ITI 231) or false, or "accepting as true" (ITI 195) and rejecting as false. I argued in the preceding chapter, however, that truth is internally related to warrant, rational acceptability and the like. And I suggested above, in partial agreement with the relativist, that understanding is also related to justification and warrant and does not require translation. Since the possibility of understanding opens up the possibility of misunderstanding, and since interpretation is required to achieve understanding where there has been misunderstanding or a lack of understanding (as on first hearing another language), interpretation turns out to be internally related to justification and rationality. This is all that we need to reconstruct Davidson's argument.

What appeals, such as we have considered, to warranted assertibility and related ideas show, I think, is that the principle of charity, as it is stated, commits us to more than we need when we go interpreting, and hence, to more than we are justified in assuming. This is the force of Putnam's and Hacking's remarks cited near the beginning of this section. Granted, I must continually balance belief-attribution with assignments of meaning, raising hypotheses and testing them, when I set out to translate the sentences of another or to interpret that person's behaviour, verbal or non-verbal. But I need not assume that I share with the speaker as many beliefs as Davidson would have us suppose. Hence, I may well attribute what I suppose to be *false* beliefs to the speaker; what I cannot do, if interpretation is to get off the ground, is attribute to the speaker a system of beliefs that I would take to be wildly incoherent, explicitly and systematically contradictory, etc. Nor can I attribute beliefs and meanings to her that seem to have no intelligible relation to her practical activity. I must assign meanings and beliefs to her in such a way as to *make sense* of what she says and what she does, and I must stand prepared to *modify* my conception of what "makes sense" as I come to understand more about the culture and situation of the person whose behaviour I

---

[62] This suggests the possibility of different conceptual schemes that are practically--but not logically--incommensurable, a point to which I shall return in §V.

am interpreting; I must, in Putnam's words, "maximize the humanity of the person being interpreted" (RTH 117).[63]

It is futile to object that, after following such a strategy to the best of our abilities, we might yet have failed systematically to understand the speaker whose language confronts us--or she, us. It is futile, because ultimately it is incoherent. Interpretation is not an all-or-nothing affair. There is always room for refinement and for misunderstanding, but never room for systematic delusion--provided we adhere to the principle of humanity in interpretation.

Some of Kuhn's later comments on his own position fit well with the revision of Davidson's argument that I have proposed, though his use of "translation" is reminiscent of Davidson's. In the 1969-Postscript to *The Structure of Scientific Revolutions* he seems to suggest that neither a complete nor a partial failure of communicability separates the practitioners of disparate paradigms; it is just a difficulty of translation that results in their disputes. As he goes on to clarify, scientists can come to understand what it is that they disagree about by trying to "translate" from one paradigm to another:

> [W]hat the participants in a communication breakdown can do is recognize each other as members of different language communities and then become translators. ... Having isolated ... areas of difficulty in scientific communication, they can next resort to their shared everyday vocabularies in an effort further to elucidate their troubles.[64]

This passage appears to provide us with yet a third Kuhn--or perhaps the first two were merely impostors in Kuhn-skin caps.[65]

Indeed, the shift from the principle of charity to the principle of humanity seems to capture better the view that Davidson himself now espouses. In "The Myth of the Subjective" we find that for Davidson incommensurability is no longer equatable with a failure of inter-translatability. It now keeps company with "unintelligibility" and a

---

[63] This preference for the principle of humanity over the principle of charity is similar, I think, to Rorty's stress on "hermeneutics." See PMN 315-394. For criticism of the alleged differences between "charity" and "humanity" see Malpas (1992), pp.154-159. Malpas claims that Davidson's principle of charity allows for those things that the principle of humanity allegedly adds to it. This question is closely connected with the question of how we should treat Davidson's uses of 'truth' and 'interpretation'.

[64] Kuhn (1969), p.202.

[65] This invites us to speculate on the extent to which the position that we have arrived at can accommodate Kuhn's account of how scientific revolutions occur, without lapsing into relativism. Putnam compares his own views and those of the "real" Kuhn:

> Kuhn has rejected this [relativist] interpretation of [*The Structure of Scientific Revolutions*] and has since introduced a notion of 'non-paradigmatic rationality' which may be closely related to if not the same as what I [have] called 'non-criterial rationality'. (RTH 113)

permanent transcendence of "rational resolve." What Davidson rejects is "conceptual relativism"--"If by conceptual relativism," that is,

> we mean the idea that conceptual schemes and moral systems, or the languages associated with them, can differ massively--to the extent of being mutually unintelligible or incommensurable, or forever beyond rational resolve ...[66]

The argument from interpretive charity shows that truth and rationality are not relative to historical eras or to cultures. But that does not imply that truth and rationality can therefore be given explicit formulations that would allow us to determine of any particular belief whether it were justified or not, true or not. We do, of course, want to make judgments about what we ought to believe, and justification is crucial when it comes to any claim that someone's behaviour ought to be affected or regulated by the truth or falsehood of certain propositions. But our standards for making such judgments are themselves open to criticism in a way that makes possible their own extension. It is this open-ended, revisable rationality that we presuppose whenever we set out to interpret the behaviour of another, and it is likewise this rationality which is flexible enough to acknowledge Kuhn's history of science without embracing relativism. We have our own conceptions of what is rational, our own "styles of reasoning" as Hacking puts it,[67] more or less local to our culture and era, and other cultures and eras have their own conceptions of what is rational, but what is presupposed by such talk is that there is some concept of which we have differing conceptions.[68] When I try to understand another person, of my culture and era (or of others), I presuppose that I can make some sense of her behaviour, linguistic and non-linguistic. But that presupposition, as trivial as it may seem, is just the presupposition that rationality transcends cultures and eras, even if we cannot say exactly what rationality is.

IV. Translation and Indeterminacy

Shifting the focus of Davidson's argument from translation and truth to interpretation and warrant suggests the relevance of this argument for the views of a respected conceptual-schemer, W. V. Quine. Quine's "pragmatism," we saw in Section I, was "more thorough" than Carnap's and did not "leave[ ] off at the imagined boundary between the analytic and the synthetic."[69] For Carnap there was no cognitive ground for choosing between theories or "linguistic frameworks," equally well confirmed by the same evidence. For Quine, the evidence

---

[66] Davidson (1989), p.159f.

[67] See Hacking (1982). I noted in earlier chapters that Hacking's criticisms of Putnam seem misplaced, because Putnam is in fact advocating a position rather like that from which Hacking draws to make those criticisms. Hacking's "styles of reasoning" seem to me to be much the same as Putnam's "conceptions of rationality."

[68] See RTH 116f.

[69] Quine (1980), p.46.

is itself internal to the theory, since there is no boundary between theory and observation. Accordingly, the non-cognitive nature of theory-choice extends to the level of entire conceptual schemes. This is evident in Quine's doctrine of the indeterminacy of translation.

In radical translation we confront a speaker of a "remote language" and, "unaided by prior dictionaries," try to translate her utterances "on behavioural evidence."[70] The final end is to give a translation-manual for her language. But Quine holds that determinate translation of another language into one's own is possible only for certain parts of that language--its "observation sentences" and its terms for "Yes" and "No." For the rest of the language there are indefinitely many systems of "analytical hypotheses" (WO 68) that can be given, each of which will be equally confirmed by the behavioural evidence given by a speaker in the situation of radical translation. For huge portions of the language no "genuine hypotheses" about translation can be formed.

To start, Quine suggests, we must pick something observable in our environment and, assuming that it will also catch the eye of the speaker and that the speaker will be prompted to make some utterance in response to this stimulation, formulate a hypothesis about how the speaker's utterance is to be translated. By repeating the speaker's utterance we may then try to elicit further utterances that may be provisionally translated as roughly "Yes" or "No."

Although for Quine there is no firm difference between statements of theory and of observation, he does admit a scale of observationality, according to which "occasion sentences" are ones to (or from) which a speaker will assent (or dissent), depending greatly on the occasion of their utterance. They "command assent or dissent only if queried after an appropriate prompting stimulation" (WO 35f). Those occasion-sentences that are tied most completely to specific stimulus-conditions Quine calls "observation sentences" (WO 42). The sentence 'Red' has greater observationality than, *e.g.*, 'Rabbit', says Quine, since "there is less scope for collateral information in deciding whether a glimpsed thing is red than in deciding whether it is a rabbit" (WO 40f). The higher the degree of observationality, the less vulnerable is a sentence's stimulus-meaning to confusion with other sensory cues. For observation-sentences, Quine holds, it makes good sense to treat stimulus-meaning as meaning-proper.[71]

It is when we try to move beyond observation-sentences that we reach the edge of the grimpen of indeterminacy with no secure foothold. The problem arises when we try to parse observation-sentences into *words*

---

[70] Quine (1969), p.45.

[71] Occasion sentences whose stimulus meanings vary none under the influence of collateral information may naturally be called *observation sentences*, and their stimulus meanings may without fear of contradiction be said to do full justice to their meanings. These are the occasion sentences that wear their meanings on their sleeves. (WO 42)

By "stimulus meaning" we are to understand the ordered pair of a sentence's affirmative stimulus-meaning ("the class of stimulations" (WO 32) prompting assent) and its negative stimulus-meaning (the class of stimulations prompting dissent), where both are relative to "the modulus, or maximum duration recognized for stimulations" (WO 33).

that can then appear in other contexts. Here the ties with occasional stimulus-conditions are broken. If we find ostensibly elliptical constructions like "Rabbit," we might be led to think that an account of the stimulus-meaning of the *word* 'rabbit' is also at hand. But we would be mistaken. It is one thing to interpret a speaker's utterance as assenting to the sentence 'Gavagai', which we might then translate as 'Rabbit'. It is quite another to think that the *word* 'gavagai' has the same referent as the English *word* 'rabbit'. The latter assumption tacitly attributes to the speaker the ontology of the translator. But it is possible that, while we find it natural to talk of rabbits, the speaker prefers to talk of incarnations of the rabbit-god or "sundry undetached parts of rabbits" or temporal "rabbit stages" (WO 52).

This "inscrutability of reference" as Quine calls it,[72] is not to be settled by the likes of stimulus-meaning. Suppose that we introduce a system of "analytical hypotheses" (WO 68)[73] that would let us isolate sufficiently short parts of speech as words and identify likely logical particles, plural endings, inflections and other elements of syntax. To the extent that these analytical hypotheses seemed to make sense of the speech-behaviour of our speaker, and of her community at large, and to the extent that we were able to interact linguistically with the speaker's community in a way that allowed us to suppose reasonable differences of belief when we encountered difficulties, we might consider our translation successful. And, indeed, it would be for practical purposes--but we would not, Quine argues, have escaped the indeterminacy that faced us, for it would remain open to us to start all over again with a different set of analytical hypotheses.

How would we tell whether 'gavagai' referred to undetached rabbit parts or to rabbits? We might ask whether one gavagai was the same as another, pointing to one part of the rabbit and then to another. But

> ... if one workable overall system of analytical hypotheses provides for translating a given native expression into "is the same as," perhaps another equally workable but systematically different system would translate that native expression rather into something like "belongs with." Then when in the native language we try to ask "Is this *gavagai* the same as that?" we could as well be asking "Does this *gavagai* belong with that?" Insofar, the native's assent is no objective evidence for translating "gavagai" as "rabbit" rather than "undetached rabbit part" or "rabbit stage."[74]

Exactly the same evidence that confirms one set of analytical hypotheses would equally well confirm other sets--indeed, there seems to be no upper bound on the number of different such sets we could propose, depending only on how clever we were and what beliefs we were willing to attribute to our speaker. But if so, then translation seems indeterminate; there is no right way of getting at the relatively non-observational utterances of a speaker and no right way of parsing her speech into words. "[I]t would be forever impossible to know of one of

---

[72] See Quine (1969), p.37.

[73] See *ibid.*, p.33.

[74] *Ibid.*

these translations," says Quine, "that it was the right one ..."[75]-- impossible, for there would be no fact of the matter.

The indeterminacy of translation is not for Quine a standard case of the underdetermination of theory by data. Even in his most recent substantial work he still holds that although "There is an evident parallel between the empirical underdetermination of global science and the indeterminacy of translation," it remains the case that

> ... the indeterminacy of translation is additional to the other. If we settle upon one of the empirically equivalent systems of the world, however arbitrarily, we still have within it the indeterminacy of translation.[76]

Whence comes this asymmetry between "global science" and the inquiries of the field linguist?

It is an incoherence, I think, to hold that I must form hypotheses about the meanings of my own words in order to understand them and a worse incoherence to suppose that there might be no fact of the matter about those meanings. But Quine suggests that there is no principled difference, either between interpreting a speaker of another language and interpreting a speaker of my own language, or between either of these cases and the case of interpreting myself.[77] Thus, if the indeterminacy of translation threatens the inquiries of the field linguist, it also threatens to creep into my understanding of myself.

To reject this threat is not to deny that I can be mistaken about myself, or my intentions; nor is it to deny the need for interpretive hypotheses where understanding fails. And this rejection does not alleviate the problem of indeterminacy only in my own case. Quine's indeterminacy-thesis violates a constraint that we encountered in Section III: *it presupposes a conception of warranted assertibility independently of a conception of interpretation.* Quine's privileging of the notions of assent and dissent is tantamount to the assumption that we can understand warrant without understanding the assertion that is warranted. Terms for 'Yes' and 'No', terms indicating the speaker's acceptance of a sentence as warranted, are treated by Quine as being available to the field-linguist in a way that other words of the language under examination are not.[78] Assent and dissent fall into the class of expressions whose determinate translation is more or less possible, whereas the greater part of the language can be dealt with only by some set of analytical hypotheses, where many other sets of such hypotheses would do just as well. Even without interpreting the rest of the language, we could still determine terms for assent and

---

[75] *Ibid.*, p.29.

[76] Quine (1990), p.101.

[77] See Quine (1969), p.46.

[78] Of course, Quine would object to my assertion that 'Yes' and 'No' have something to do with warrant, even if they are linked with the attitudes of "holding true" and "holding false." For Quine truth remains non-epistemic (see *ibid.*, p.93f). I have, however, already advanced arguments for the contrary view. See Chapter 2, §§V-VI.

dissent; we could still make sense of assertibility.[79] The bulk of another language might be inaccessible to us--or worse--there might be *no fact of the matter* concerning how to interpret most of the language; yet, within the range of underdetermination that Quine takes to characterize physics, we could recognize certain utterances as warrantedly assertible in that language.[80]

Of course, Quine might demur from this description of his position and assert the pervasiveness of his holism. To do so would be more consistent than to accept the asymmetry of observation sentences, assent and dissent, on one hand, and everything else, on the other. But it would also undermine the indeterminacy of translation.[81] As Rorty says, Davidson's position goes beyond Quine's by recognizing that the field-linguist "must be purely coherentist in his approach, going round and round the hermeneutic circle until he begins to feel at home."[82] By denying this recognition, Quine replaces Carnap's dualism of the synthetic and the analytic with his own dualism of the observational and the non-observational--a dualism offered as a fluid gradation which then hardens in its philosophical mold.

My argument resembles Donald Hockney's critique of Quine.[83] Hockney argues that our hypotheses about assent and dissent in the speaker's language ("A-D hypotheses")[84] seem not to differ importantly from the analytical hypotheses that we ought not to take as genuine hypotheses, according to Quine. So, if systematic change of analytical hypotheses is possible, it is equally possible that we could systematically alter our A-D hypotheses, provided we made compensating changes in our "genuine" hypotheses about stimulus-meanings.

Moreover, as Hockney observes, Quine's insistence that physical theory is more determinate than linguistics is the mark of a dogmatic scientism. Indeed, this "bifurcation thesis" is inconsistent with the thesis of translational indeterminacy. The bifurcation thesis implies that hypotheses of physics are *genuine* hypotheses. However, this cannot be maintained if analytical hypotheses of translation are *not*

---

[79] This is not to say that terms for assent and dissent are self-identifying. As Quine observes, selecting one term as 'Yes' and another as 'No' constitutes "a working hypothesis" (WO 30). But, in the language I shall use below, it is a "genuine" hypothesis. A radical indeterminacy does not afflict these terms in Quine's scheme.

[80] Whether Davidson would be willing to accept this critique of Quine is unclear to me. He sometimes says that Quine's worries about the indeterminacy of translation are well-founded, but that "the range of indeterminacy is less than Quine thinks it is" (ITI 228).

[81] Thus, I disagree with a central claim of a recent book on Davidson's philosophy. See Malpas (1992). Malpas holds that Davidson's rejection of the Quinean focus on occasion-sentences and stimulus-meanings actually increases the degree of indeterminacy for Davidson (see *ibid.*, pp.34-43). For a brief critique see Hymers (1993).

[82] Rorty (1991a), p.133.

[83] Hockney (1975).

[84] *Ibid.*, p.421.

genuine hypotheses. For if analytical hypotheses have no truth-values, then there is no fact of the matter regarding the meanings of relatively non-observational sentences of a given language. And among those non-observational sentences will be most of the claims of theoretical physics. In Hockney's words,

> [A] sentence has a truth value only if some statement asserting that it means such and such is true. Thus theoretical statements of physics are neither true nor false. They are not genuine hypotheses.[85]

Perhaps such a conclusion might be avoided, were it open to Quine to claim that indeterminacy is a problem only for the case of *radical* translation, and not for interpreting one's own language. But Quine himself observes that "radical translation begins at home."[86] There is no reason in principle why the meanings of words of my own language--uttered by another or by myself--should be safe from indeterminacy, if the words of another language are not.

There is, then, no difference in principle between the field-linguist's attempts to understand an unfamiliar language, without benefit of a translation-manual, and my own self-understanding. So, Quine's indeterminacy-thesis entails the absurd consequence that I could be systematically ignorant of the meanings of my own words--not because they had some meaning of which I was deluded, but because there was no fact of the matter regarding their meanings. But the consequence that I might not understand my own words is one we have met before. It was a possibility for the causal theorist's account of reference that I might be systematically deluded about the reference of my words without my delusion's manifesting itself in any observable way. Similarly, metaphysical realism threatened us with the absurd consequence that we might not know the reference or meaning of our words, by allowing the real possibility of our being brains in a vat.

The recurrence of this absurdity is no coincidence, for the assumptions leading to it are at root one. The problem for the causal theorist, we saw, lay in the fact that she treats reference as a *non-intentional* relation. Similarly, the metaphysical realist made the mistake of treating possibility as being radically *non-epistemic*. Quine is at home in this tradition of thought with his "epistemology naturalized"[87] and his assumption that truth is non-epistemic.[88] When Quine treats assent and dissent as notions that are discernible in another's language independently of our having interpreted the rest of the language, he treats these notions as detached from *meaning* and *warrant* and as identifiable with the intratheoretical exactness of any *physical* relation. Similarly, when the causal theorist supposes that reference is a non-intentional relation, she supposes that truth can be understood apart from any particular interpretation of a language. When the metaphysical realist says that Cartesian scepticism is both

---

[85] *Ibid.*, p.417.

[86] Quine (1969), p.46.

[87] *Ibid.*, pp.69-90.

[88] See Quine (1990), p.93f.

unanswerable and uninteresting, she implies that the truth of a sentence can be evaluated without our understanding its meaning.

In rejecting Quine's indeterminacy-thesis, then, I am merely being consistent with my earlier rejection of causal theories of reference and metaphysical realism. I reject a further dogma--perhaps not simply of empiricism--namely, the dogma of scientism, which manifests itself here as an attempt to replace the normative with the nomological.

V. Interests, Activities and Incommensurability

I began with Davidson's assumption that the incommensurability of conceptual schemes should be understood in terms of translatability. And although we have seen that the possibilities relevant to this controversy are ones of interpretation and understanding, not ones of translation, I have maintained that an argument parallel to Davidson's serves to show that sense cannot be made of conceptual relativism in hermeneutical terms either. However, should incommensurability be linked so closely to questions of translation or interpretation? Code, as we saw, expresses reservations about the import of concluding that there is no *logical* barrier to our understanding other cultures, when there are clearly practical barriers that are not removed by defeating sceptical worries. Charles Taylor writes of incommensurable activities that are "incompatible in principle"[89] and of "incommensurable ways of life."[90] And Paul Feyerabend criticizes Putnam for attributing to him a belief in incommensurability in the Davidsonian sense.[91] Is there room for some other account of incommensurability?

Much of the plausibility of the thesis of incommensurability in the context of language and translation stems from an inclination toward a needlessly formalistic view of language. Faithful to the legacies of philosophers of language like the author of the *Tractatus*, and of linguists as diverse as Saussure and Chomsky, some thinkers suppose that a language is an abstract entity that can be usefully considered independently of any particular contexts of use--*la langue* is the immanent spirit of *la parole*; we understand a language by considering its deep structure, not the way in which its components are used. But this is to forget that, as Malpas remarked earlier, "Interpretation is ... always a practical task oriented towards the particular, rather than a theoretical one oriented towards the universal."[92] Languages are always in flux, and nothing in the present context is to be gained from ignoring this fact. We cannot regard languages as embodying incommensurable conceptual schemes in Davidson's sense, because there is no formal structure that is the language of a culture, and so, we cannot hold up such discrete structures--in imagination or otherwise--and conclude that there is no point of comparison. Davidson seems to have this point in mind in a recent paper, when he writes, "I conclude

---

[89] Taylor (1982), p.98.

[90] *Ibid.*, p.100.

[91] See FR 272.

[92] Malpas (1992), p.112.

that there is no such thing as a language, not if a language is anything like what many philosophers and linguists have supposed."[93]

Rorty gives a similar critique of formalism in his essay "Solidarity or Objectivity?"

> [A]lternative cultures are not to be thought of on the model of alternative geometries. Alternative geometries are irreconcilable because they have axiomatic structures, and contradictory axioms. They are *designed* to be irreconcilable. Cultures are not so designed, and do not have axiomatic structures.[94]

Our only way of comparing our language to another is through using it, and the process of comparison in this case automatically alters the things being compared. Comparing our language to another constitutes an *extension* of our language, like any other use of words. This is part of Feyerabend's response to a criticism from Putnam:

> [I]ncommensurability is a difficulty for philosophers, not for scientists. Philosophers insist on stability of meaning throughout an argument while scientists, being aware that 'speaking a language or explaining a situation means both *following* rules and *changing* them' ... are experts in the art of arguing across lines which philosophers regard as insuperable boundaries of discourse. (FR 272)[95]

A development of our language is the natural product of an attempt at dialogue and mutual understanding, what Feyerabend calls an "open exchange" (FR 29n13). Nonetheless, prior to any meeting of cultures or theories or languages it may well be true that there are terms in one language that are not translatable into the other. It is only with time that each culture acquires additions to its vocabulary, perhaps borrowed from the other culture and accompanied by a consciousness of associated practices, and then it may become possible to translate--and maybe not. It is not the case, to use Feyerabend's formulation again, "that English *as spoken independently of the comparison* already contains native [*i.e.* other-cultural] ideas." Rather, "languages can be *bent* in many directions and ... understanding does not depend on any particular set of rules."[96] It remains a possibility for us to come to understand another language and culture, to learn to interpret its language, even if we cannot adequately translate portions of it, because understanding need not be a matter of translation.[97] We can come to understand by a judicious mix of explicit interpretation and--

---

[93] Davidson (1986), p.446. It is tempting to see this paper as a kind of *reductio* of the philosophical interest of Davidson's long-standing attachment to Tarski truth-definitions.

[94] Rorty (1991a), p.26.

[95] Mary Hesse makes much the same point. See STSS 38.

[96] Feyerabend (1988), p.197.

[97] See MacIntyre (1987), p.393 for similar remarks.

if we are allowed--participation in the activities that give meaning to the language whose speakers we are trying to comprehend.

These considerations help to make clear that the incommensurability of conceptual schemes is at its root a *practical* incommensurability to be dealt with in a practical manner. I follow Davidson in rejecting the "third dogma" of empiricism, but I do so, not by saying that the notion of a conceptual scheme is empty, but by saying that there is no clear distinction to be drawn between a scheme and its content, nor between one scheme and another. Scheme and content are, as physicists say, two different "degrees of freedom," components that together form the vector of a language--but there is no absolute space, and hence, no absolute distinction between the *x*-axis and the *y*- or *z*- axes. A conceptual scheme, seen aright, is not a formal, eternal structure--a synchronic time-slice of our language made general and implicitly diachronic. It is a motley assortment of related (but different) practices that can grow and change to embrace new practices. Nonetheless, schemes are amalgams of historical customs that may differ from one another--hence, they embody practical differences.

Moreover, practical differences are precisely what matter,[98] because they leave room for a kind of incommensurability that is not touched upon by Davidson's construal of that term as tied to clashes between purely formal conceptual schemes. There can be no reasonable doubt that different cultures--different groups and individuals within a single culture--do things differently. Indeed, if there were no difference at a practical level, if people merely said different sorts of things while going about their business in a largely homogeneous way, the whole issue of conceptual relativism would seem arcane. (Some might say that it already is arcane--but it would be even more so.) It is hard even to imagine a difference of "conceptual schemes" in the absence of a difference of custom and practice, given the intimate connection between speaking and acting, given that speaking is a kind --or many kinds--of acting, "practical consciousness, existing also for other people" (MEW III 30),[99] as Marx and Engels say.

Wittgenstein makes the point forcefully, returning us to the kinds of considerations that motivate Davidson's views. Not only would the hypothesis of conceptual differences seem arcane in the absence of practical differences, but we could not in the worst case even recognize another culture (with whose conceptual scheme ours was said to be incommensurable) as a language-using culture. "Let us imagine," says Wittgenstein, that the inhabitants of an unknown country

> ... carried on the usual human activities and in the course of them employed, apparently, an articulate language. If we watch their behaviour we find it intelligible, it seems 'logical'. But when we try to learn their language we find it impossible to do so. For there is no regular connexion between what they say, the sounds they make, and their actions; but still these sounds are not superfluous, for if we gag one of the people, it has the same consequences as with us; without the sounds their actions fall into confusion ... (PI §207)

---

[98] "[A]ny theoretical confrontation must also be a practical confrontation if it matters at all" (Lovibond, 1983, p.153).

[99] See Marx and Engels (1970), p.51.

The ethics of gagging a subject under study may be dubious, but Wittgenstein's point is clear enough: it is inappropriate for us to think of these people as speakers of a language. "There is not enough regularity for us to call it 'language'" (PI §207). There is not enough regularity, because language is part of practical human activity and cannot be understood apart from such activity. To suppose that a people's behaviour might display the complexity and uniformity of a culture of language-users, but that in this one respect they might differ--their language is incomprehensible to us--is just to suppose that they do not really speak a language. To just the extent that we can understand the behaviour of these people, we can also understand their language. The situation that Wittgenstein describes is one in which the perfect appearance of a language is not supported by any reality, in which the incommensurability of conceptual schemes slips into scepticism about other minds.

If we are to make sense of incommensurability, then we must think of it as rooted in incommensurable practices or activities, as Taylor suggests, and when we consider such incommensurabilities, it becomes apparent both that they are *possible*--cultures and sub-cultures do come into practical conflict--and that they are *limited* in scope. Whatever differences may obtain within a culture or from one culture to the next, there is also a considerable degree of commonality in practices, guaranteed by the fact that, in general, human beings require nourishment, shelter, and human contact and affection. As biological organisms we have an *interest* in these things. Indeed, it is this commonality of interests and consequent practices which, Wittgenstein suggests, "is the system of reference by means of which we interpret an unknown language" (PI §206). Should we someday encounter a species whose practices are rendered very different by their different sensory apparatuses, or whose biological composition is so different that their "basic" interests seem obscure to us, then the scope of possible incommensurability is broadened--but it remains rooted in real practical differences.

The foregoing suggests a further level on which we may speak of incommensurability, the level of interests.[100] What we take to be in our interests usually plays a role in motivating our practical activity. A being that had *no interests* at all is one, I suggest, that we could not regard as rational. Recall Putnam's Turing Test for Reference, encountered in Chapter 2. Without nerve-endings or sensory organs a Turing machine can do no more than seem to speak a language. It has no *interests* in anything, no desires to satisfy, no ideals to fulfil, no loved ones to care for or be cared for by--no "original

---

[100] When I speak of interests and suggest rough generalizations of interests, I do not have in mind a view that is in any way friendly to "the contemporary fad for sociobiology" as Taylor aptly puts it (Taylor, 1985, p.111.) It does not follow from the fact that a rough generality of such interests (food, shelter and affection) makes understanding practically possible that normative talk is reducible to survival-and-reproduction talk. Clearly, biology is not irrelevant to understanding: in order to speak "I gotta have a brain." But this is hardly a startling revelation and certainly no comfort to advocates of biological determinism. For compelling critiques of biological-determinist science see Hubbard (1983), Gross and Averill (1983) and Rose, Lewontin and Kamin (1984).

purposes"[101] in Charles Taylor's phrase. Rather, its purposes are "derivative" or "user-relative."[102] Like Douglas Adams' fictional computer "Deep Thought," if well enough engineered and properly programmed, it could spend seventy-five thousand generations searching for the answer to "the great question of Life, the Universe and Everything" and be satisfied with the result "forty-two."[103]

But while human beings share many basic ends, the means that they arrange for the attainment of those ends vary widely, and in the course of devising means we also acquire new and divergent ends. Since language is a part of our activity, it should not seem strange that different kinds of languages, suited--more or less--to different purposes, should arise, nor that the concepts found in those languages should differ as much as the purposes for which they are needed. "[A]n education quite different from ours," Wittgenstein suggests, "might also be the foundation for quite different concepts."[104]

> For here life would run on differently.--What interests us would not interest *them*. Here different concepts would no longer be unimaginable. In fact, this is the only way in which *essentially* different concepts are imaginable.[105]

Elements of the practices that serve our basic needs come to matter to us themselves, and their attainment and preservation give rise to new practices--linguistic and otherwise. To the extent that such changes take place in settings isolated from one another, different courses of development and different concepts are to be expected.

I differ from Davidson, then, in allowing the possibility of different conceptual schemes that can be practically incommensurable. This difference is a direct consequence of shifting the focus of Davidson's argument from (a) translation-manuals and "total theories" to local, but always revisable and expandable, interpretation, bound up with the practical activity and interests that give language meaning, and from (b) the idea of truth to warrant, justification, rationality and similar concepts, which, I have argued, are internally related to truth, understanding and interpretation. And while my modified version of Davidson's argument against abstract conceptual schemes does not rule out the practical incommensurability of the actual concepts employed by different cultures, none of this entails the unintelligibility of other concepts, languages, cultures or persons. Understanding, as we have seen, is not reducible to translation, and rationality is open-ended and revisable.

---

[101] Taylor (1985), p.99.

[102] *Ibid.*, p.99.

[103] Adams (1979).

[104] Wittgenstein (1981), §387.

[105] *Ibid.*, §388.

Chapter 4: Relativism, Truth and Rationality

In the preceding chapter I argued against "conceptual relativism," the view that different groups could possess different conceptual schemes such that the members of one group could not in principle understand the members of the other group. This view, I argued, is a disguised form of metaphysical realism, not with respect to the "external" world, but with respect to conceptual schemes. In casting doubt on our abilities to understand the members of other cultures the conceptual relativist raises a kind of scepticism about other minds, dressed in cultural clothing. Conceptual relativism in this sense suffers from an epistemic neurosis comparable to that which afflicts metaphysical realism, for the conceptual relativist does not really doubt the existence of other-cultural minds. Still, her view depends on the real possibility that just such a doubt might prove correct.

Despite its problems, this form of relativism is, I think, the most plausible, *prima facie*. In this chapter I shall turn to a less plausible kind of relativism--relativism about truth and rationality, or "epistemic relativism." We have already seen that, in response to Davidson's initial argument against the idea of a conceptual scheme, the conceptual relativist may be tempted to turn to epistemic relativism to avoid the contention that we have no conception of truth independently of our conception of translation. Truth, says the relativist, reduces to warranted assertibility (or an idealization thereof) where what constitutes warrant varies from one culture to another. Truth, too, then varies from one culture to another. This move does not save conceptual relativism, I have argued, because a revised version of Davidson's argument cites the link between warrant and interpretation. We have no notion of warranted assertibility independently of the notions of interpretation and understanding. The dedicated relativist, however, might try to hold that it remains the case that truth and rationality are relative to cultures. So, we must investigate the coherence and justifiability of *epistemic* relativism.

I shall examine some traditional reasons for accepting such relativism, arguing that they are inconclusive. The sorts of reasons that I shall consider are, for the most part, arguments against metaphysical realism--much like those I have endorsed. Such arguments do not lead to epistemic relativism. Inconclusiveness is the least of the relativist's problems, however, for the position that she endorses is at its worst self-defeating and at its best question-begging. To argue for her position the epistemic relativist must presuppose a non-relativistic conception of truth, and so, finds herself troubled by the compulsions of another epistemic neurosis. Although she claims to doubt non-relativistic accounts of truth, she cannot avoid using one if she is to justify her position. The relativist gets into difficulty by granting the assumption that truth must be independent of warrant, or "radically non-epistemic,"[1] if it is to be non-relative.

---

[1] Putnam (1978), p.125.

I. Relativism as a Response to Realism

The arguments that I shall focus on here depend for their force on assuming the dichotomy of realism and relativism and then showing realism to be incoherent. One of the most committed endorsements of this strategy in recent years has come from proponents of the "Strong Thesis" (ST) in the sociology of knowledge. Advocates of the Strong Thesis hold that "true belief and rationality are just as much explananda of the sociology of knowledge as error and non-rationality, and hence that science and logic are to be included in the total programme" (STSS 31f). As Mary Hesse indicates, ST can be understood in a way that is "true but merely trite" (STSS 32). Allowing that the genesis of beliefs that are held to be true or justified is worthy of sociological inquiry does not entail relativism, because even if we reject relativism we can still ask for a causal story about the biological or sociological mechanisms implicated in a person's holding a particular belief. Such a causal story might be correct

> ... even if it were also true that there are conclusive arguments for the necessity of rational rules, for it is clear that few people (if any) in fact adopt such rules on pure grounds of rational necessity uncaused by any previous social history. (STSS 32)

But what is intended by the Strong Thesis is some postulate of "symmetry" or "equivalence" (RRSK 22),[2] which entails that all thinking is historically and culturally situated in such a way that "even if there are rational rules that are ultimately independent of social (and perhaps biological) causation, at least no such rules can be appealed to as independent variables in a social explanation of knowledge" (STSS 32). The postulate is one of "symmetry," because, without access to any "extra-natural and extra-social grounds of rationality and truth in the *a priori*, the analytic, or the necessary" (STSS 37), beliefs can be assessed only relative to what a culture holds rational or true. Hence, "all beliefs are on a par with one another with respect to the causes of their credibility" (RRSK 23). We have no access to necessary truths that would command assent on pain of irrationality; all that we have are our own culturally relative standards, comprising one set of standards among many. Two cultures may disagree about the truth or rationality of a belief, but if so, there is no way of adjudicating between their pronouncements, because assessments of rationality can be made only by reference to the same culturally relative criteria that brought about the conflict in the first place (or by reference to some third culture's similarly relative standards). If we are to understand why a particular belief is held, it will be useless to classify it as rational or irrational. Instead, the sociologist of knowledge must treat "rational" and "irrational" beliefs, "true" and "false" beliefs as equivalent with respect to their credibility and inquire, rather, into the social, psychological or biological causes of that credibility. So say advocates of the Strong Thesis.

There is, I think, a mistake in the reasoning of the ST-advocate, and it is a mistake that arises from an acceptance of the dichotomous

---

[2] See STSS 32.

opposition of realism and relativism. Hesse seems aware that problems can arise from this dichotomy, when she writes, "it is often an unreasoning fear of relativism that goes far to explain the power with which metaphysics seems to have caught the imagination of realist philosophers."[3] Indeed, as we shall see shortly, Hesse recommends that the claims of the Strong Thesis be weakened, though I do not think she succeeds in rescuing even this less forceful version. Barnes and Bloor, however, despite subscribing to a number of plausible claims about language and meaning, seem less aware that a rejection of metaphysical realism does not leave us in the embrace of relativism. I shall now establish the plausibility of this charge.

In my summary of the position represented by the Strong Thesis I alluded to an argument *from* the historically and culturally situated nature of judgments of truth and rationality *to* the conclusion of epistemic relativism. Barnes and Bloor make the argument as follows:

> For the relativist there is no sense attached to the idea that some standards or beliefs are really rational as distinct from merely locally accepted as such. Because he thinks that there are no context-free or super-cultural norms of rationality he does not see rationally and irrationally held beliefs as making up two distinct and qualitatively different classes of thing. They do not fall into two different natural kinds which make different sorts of appeal to the human mind, or stand in a different relationship to reality, or depend for their credibility on different patterns of social organization. Hence the relativist conclusion that they are to be explained in the same way. (RRSK 27f)

Much the same argument is at work a few paragraphs later where they insist that "Validity totally detached from credibility is nothing" (RRSK 29), and that since we have no access to necessary truths, the metaphysical realist, despite contrary pretensions, will also "treat validity and credibility as one thing by finding a certain class of reasons that are alleged to carry their own credibility with them: they will be visible because they glow by their own light" (RRSK 29).

I have fewer problems with the premises on which Barnes and Bloor rest their argument than with their conclusion,[4] but the premises are not unimpugnable. It is one thing to say that rationally and irrationally held beliefs "do not fall into two different natural kinds" and quite another to say that they do not differ at all. It is

---

[3] Hesse (1980), p.xiv.

[4] For an explicit acknowledgment of Wittgenstein's influence see Bloor (1976), Bloor (1983) and Barnes (1987). See also RRSK 37f. My differences with them extend to the interpretation of Wittgenstein, whose work they treat, I think, too much as though it were Quine's. This is especially clear in Barnes (1987), where arguments similar to those of the rule-sceptic are used to try to show that

> Where [empirical] assertions use terms learned wholly by ostension it would seem that their truth or falsity must be indeterminate, since the future application of such terms is not fixed in advance. (Barnes, 1987, p.25)

one thing to say that "validity totally detached from credibility is nothing" and quite another to say that they are "one thing." It is one thing to say that our judgments of validity and credibility are historically and culturally situated and quite another to say that there is no distinction between a belief's being rational and its being "merely locally accepted as such." Barnes and Bloor, I submit, unjustifiably conflate three things with three other things. Or more accurately, they conflate two things in these three different ways.

Consider the second manner of conflation. The idea that I endorsed in Chapter 2, that truth is internally linked with idealized warranted assertibility, is just the idea that the distinction between validity and credibility is "not an absolute distinction" (RRSK 30). However, it does not follow from the rejection of an absolute distinction that we are left with no distinction at all. The metaphysical realist, we saw, takes truth to be "radically non-epistemic," especially if she subscribes to a correspondence-theory of truth. The problem for the correspondence-theorist, I argued, following Putnam,[5] is that there are indefinitely many correspondences between words and the world of the metaphysical realist and no non-arbitrary way of saying which of these is *the* truth-making correspondence. The source of the problem is that the correspondence-theory treats truth as a non-intentional or non-normative phenomenon. In other words, it drives a wedge between credibility and validity. However, acknowledging that truth has an epistemic component is compatible with preserving the time-honoured distinction between being right and thinking that I am right. Truth is an ideal that regulates assessments of warrant, and we can maintain that truth is tied to an idealization of warrant without holding that it is *reducible* to warrant. A belief can be warranted without being true, and this is implicit in the very notions of truth and warrant.

Now, what of the first conflation? While rationally and irrationally held beliefs do not fall into two *natural* kinds, they certainly do fall into two *normative* kinds--*viz.*, the rational and the irrational. And the fact that no set of conditions is necessary and sufficient to distinguish the rational from the irrational is no barrier to *our* distinguishing them. No set of necessary and sufficient conditions distinguishes my desk from the "unoccupied" space adjacent to it--at the molecular level the one fades into the other without demarcation--but the one is clearly not to be confused with the other. Perhaps the boundaries of physical objects will seem too dissimilar from the boundaries of concepts. Then consider the distinction between the literal and the metaphorical. No set of necessary and sufficient conditions distinguishes a metaphorical use of an expression from a literal use of the same expression. Indeed, old metaphors often don't die; they become new paradigms of the literal. (Think of expressions like "heartless.") But the one category is not to be confused with the other. The rational and the irrational have similarly blurry edges, but this does not render their differences incomprehensible.

Furthermore, that the rational and the irrational are two distinct kinds--even if no list of necessary and sufficient conditions can distinguish them--follows quite "naturally" from the internal relation between truth and justification. A belief or a proposition can be true and unjustified, and a justified belief or proposition can be false, but a true belief or proposition is not *really* unjustifi*able*--*i.e.*, it

---

[5] See Chapter 1.

is not without justification in every "really possible world" (assuming the constraints placed on this latter notion in Chapter 2). A false proposition or belief might well (but need not) be really unjustifiable, and will not be justifiable under epistemically ideal conditions (*for that belief or proposition*).

Nor does the fact that "[o]ur norms and standards of warranted assertibility are historical products"[6] and, indeed, cultural products, entail that rationality or warrant is (for us) nothing but what is accepted as such by our culture. For, as I noted in Chapter 3, those norms and standards are revisable and subject to critique. Reason is capable of self-criticism--or if this sounds too metaphysical, reasonable people are capable of self-criticism, and the members of a culture or sub-culture are capable of scrutinizing and changing their standards. But being capable in this way just means that self-reflection and discussion can be more than an act of reaffirming what we already believe (though, to be sure, the latter phenomenon is not unknown). And we should not forget that insofar as cross-cultural understanding is possible--and its possibility is just the possibility of recognizing other cultures as being other cultures --discussion and criticism can take place across cultural "lines."[7]

My disagreement with ST-advocates over the significance of the historical and cultural conditioning of the standards of rationality that people actually apply, when engaged in discursive activity, is linked with the kinds of arguments that Barnes and Bloor, and Hesse, employ to establish the fact of this conditioning. Their central line of reasoning relies on Quine's claim that theory is underdetermined by data and on his related rejection of the analytic-synthetic distinction. ST-advocates take very seriously Quine's suggestion that, if analytic statements are thought of as those "which hold come what may," then "[a]ny statement can be held true come what may, if we make drastic enough adjustments elsewhere in the system"[8] that comprises the "totality of our so-called knowledge or beliefs."[9] There are no *a priori* truths or falsehoods, just statements that we in our culture at a certain point in history find convenient to treat as *a priori*.

However, I cautiously reply that the fact that we can rationally criticize standards of rationality themselves is roughly the idea that "there is at least one *a priori* truth" (PP3 98).[10] When we suggest

---

[6] *Ibid.*, p.21.

[7] Recognition of the possibility of criticizing one's own standards and practices is displayed in Barnes (1987), pp.39-42, but Barnes does not seem to appreciate the consequences of this for relativism. The reason, I think, is that he exempts interests from the possible objects of critique: "All interests do is to prompt specific choices between rationally defensible alternatives" (*ibid.*, p.29).

[8] Quine (1980), p.43.

[9] *Ibid.*, p.42.

[10] Putnam displays a candid ambivalence about this claim, but observes that, even if it is risky to try to show of any particular statement that it is an *a priori* truth, "it is not just dangerous but wrong to make the quick leap from the fact that it is dangerous to

revisions to our existing practices we posit an ideal of rationality as the limiting concept against which our present norms are to be compared. Of course, we have no way of saying infallibly just what is really *a priori* for reasons much like Quine's. Many of our most central presuppositions can be called into question or rejected, provided we are prepared to "make drastic enough adjustments elsewhere in the system." But this is quite compatible with supposing that some logically contingent propositions are ideally justified--*e.g.*, "We have not always been brains in a vat." And this sort of ideally justified, or "certain," contingent proposition ("certain" in the non-Cartesian sense of Chapter 2) seems a good candidate for something like a synthetic truth known *a priori*.[11] As well, in conceding the fallibility of assessments regarding what is *a priori* we must still keep two points before us: first, the "system" in which we may need to "make drastic ... adjustments" is not a closed system with an axiomatic structure and clearly definable rules of inference, and so, secondly, there are some tenets whose abandonment simply could not be completely compensated for by changes to our other beliefs.

This point can be illustrated by turning once again to the case of radical interpretation. Is there really *no* belief that we could not attribute to members of another culture and still regard them as language-users and *a fortiori* as thinkers? Could any circumstance ever really justify us in attributing to the members of another culture the belief that every utterance is unambiguously both true and false[12] without also justifying the conclusion that they were not really speaking a language, but were just making noises? The supposition that *any* belief could be replaced, subject to a careful retuning of the remaining ones to assure overall coherence, is just the supposition that conceptual schemes have discrete boundaries that distinguish them, one from another--an idea that we have already found incoherent. But it is an idea to which the relativist will have further recourse.

II. How Relativism Defeats Itself

The revelation that epistemic relativism tacitly relies on conceptual relativism for its credibility accords well with my suggestion in Chapter 3 that *prima facie* the most plausible route to epistemic relativism is via conceptual relativism. It also suggests that, not only are the epistemic relativist's arguments inconclusive (for the relativist's purposes), but there is a problem for the very coherence of her position. Indeed, I shall now argue that epistemic relativism is self-defeating and that it manifests much the same epistemic neurosis that I have diagnosed in the arguments of the metaphysical realist. And, as I suggested in Chapter 3, this is not

---

[10] claim that any statement is absolutely *a priori* to the absolute claim that there are no *a priori* truths" (PP3 114).

[11] Of course, if one held that "I am in pain," when honestly asserted, is ideally justified, then *this* sort of certainty would not be a plausible candidate for the synthetic *a priori*. (Thanks to Bruce Hunter.) But I shall not argue for the certainty of such avowals.

[12] This is roughly what Putnam calls the "minimal principle of contradiction" (PP3 101).

surprising, given that the relativist is ultimately *just another kind of realist*. Let me first turn to the charge of self-defeat.

The claim that epistemic relativism is self-defeating is hardly a new one. Most realists, I suspect, regard it as a truism. But I think it is worth taking the time to see why relativists persist in being relativists--*i.e.*, why they think the charge of self-defeat is just a product of the realist's reluctance to part with absolute, acultural, ahistorical standards of truth and rationality.

One reason is apparent. If metaphysical realism is an incoherent position--and I agree with the relativist that it is--and if one subscribes to the dichotomy of the objective and the subjective, then relativism is by default at least as good a position as realism from the standpoint of rationality. "Why be a relativist?" we might ask. "Well, why not?" is the reply. If coupled to this thinking is the idea--which I shall examine in Chapter 5--that epistemic relativism is a more satisfactory doctrine from a political or ethical perspective, then realism will just seem like a discourse of power to be undermined by the only tools available to do the job, the tools of relativism.

But there is another reason why the epistemic relativist remains unconvinced by charges of self-defeat; it is none other than the implicit reliance of epistemic relativism on conceptual relativism. But in order to see this, we must examine the argument for self-defeat and the replies that the supporter of relativism is tempted to make.

The reasons for thinking that relativism is self-defeating are not complicated. Truth and rationality, the relativist says, are relative to cultures and historical eras. But if so, then the relativist's thesis must itself be only relatively true, not true *simpliciter*. Thus, one can have no reason for believing it, unless one already believes it. Or, more accurately, one might have reasons for believing the relativist's thesis, but ultimately she must regard their status as reasons--their *relevance*--as itself relative.[13] A reason for being a relativist can be only a relative reason for being a relativist. On the other hand, if the relativist tries to claim that her thesis is true, not relatively true, then she has tacitly acknowledged that relativism is false. To claim that it is non-relatively true that there is no non-relative truth is to undermine that very claim. It is of a piece with the metaphysical realist's claim that the nature and existence of the world are independent of our epistemic capacities, but that Cartesian scepticism is only a logical possibility.

Now, Barnes and Bloor have little sympathy for this criticism. "The claim that relativism is 'self-refuting'," they tell us, has been "thoroughly demolished" (RRSK 23n6) by Mary Hesse (she herself is more modest). But this remark makes it apparent that Barnes and Bloor adhere to rather unusual architectural standards. Consider Hesse's version of the argument from self-defeat:

> Let P be the proposition 'All criteria of truth are relative to a local culture; hence nothing can be known to be true except in senses of "knowledge" and "truth" that are also relative to that culture.' Now if P is asserted as true, it must itself be true only in the sense of 'true' relative to a local culture (in this case ours). Hence there are no grounds

---

[13] Thanks to Bruce Hunter for suggesting this clarification.

> for asserting P (or, incidentally, for asserting its contrary). (SSTS 42)

This version of the argument, Hesse says, turns on an equivocation, because "[i]f a redefinition of cognitive terminology as relative to a local culture is presupposed in asserting P, then P must also be judged according to this redefinition" (SSTS 42). The critic, thinks Hesse, assumes that the only "grounds for asserting P" must be non-relative grounds, and so, that the relativist cannot formulate her position without violating its own demands. But, she claims, we should think of the relativist as asserting P only *relatively*, and as relying only on relative grounds to justify her position. To demand that, if P be asserted at all, it be intended as non-relatively true is simply to insist on non-relative definitions of cognitive terms.[14]

But why should we grant the redefinition of cognitive terminology presupposed by the relativist's thesis? What would this "redefinition" look like? The relativist of Hesse's account seems to say that in asserting P she assumes that cognitive terms, such as 'knowledge' and 'truth' are to be redefined as follows:

> All criteria of truth are relative to a local culture; hence nothing can be known to be true, and so, P cannot be known to be true, except in senses of 'knowledge' and 'truth' that are also relative to that culture.

But isn't this "redefinition" just a reiteration of the thesis itself? The assertion that P should be assessed by its own standards can be made with similar force about *any* proposition, with the consequence that practically every proposition is trivially justified if assessed simply by its own lights.[15] If the relativist is trying to *argue* for her position, then simply re-asserting her conclusion is not enough. By taking this route she will be begging the question.

One response to this criticism might be to say that P is an expression of a certain attitude.[16] That attitude is one of tolerance for opposing views and a reluctance either to make claims without qualifying them with the recognition that others might disagree ("I hold this to be the case, but it might not be true-for-you ...") or to accept such unqualified claims ("Be more modest! What you say may well be true for you, but it's not for me!"). However, it is difficult to distinguish this attitude from another, non-relativist attitude--the attitude of someone who is a conscientious fallibilist: "I hold this to be so, but I could be wrong, and you may well rationally disagree with me." If it is a non-cognitive construal of P that we are asked to

---

[14] Hesse herself does not seem to embrace relativism, arguing usually for a position closer to Putnam's internal realism. See, *e.g.*, the other essays in Hesse (1980). In a more recent collaborative effort the authors tell us, "We seek to avoid a relativism in which 'anything goes' ..." (Arbib and Hesse, 1986, p.10). However, I shall not offer any extended examination of that attempt.

[15] Exceptions here might include self-referential paradoxes like "This sentence is false."

[16] Thanks to Bruce Hunter for raising this.

accept, then the attitude in question is one that will--we might expect--issue forth in certain kinds of behaviour, other things being equal. If we cannot distinguish the behaviour of the relativist from the behaviour of the conscientious fallibilist--and I submit that the two attitudes do not differ significantly in the behaviour that they produce in individuals who take them seriously[17]--then there is no good reason to believe that the attitude of relativism is in any way superior to that of fallibilism. (These attitudes will be an implicit theme of much of my discussion in Chapter 5.)

But Hesse does not take the non-cognitivist route of the preceding paragraph. Rather, conceding that a *petitio principii* is involved in arguing for the truth of relativism, she urges that the redefinition of cognitive terms be taken as neither a question-begging stipulation, nor "a demonstrable conclusion from acceptable premises ..."

> ... but rather as a hypothesis in the light of which we decide to view knowledge, and consider whether its consequences are consistent with the rest of what we wish to affirm about knowledge, and whether it does in the end provide a more adequate and plausible account than the various rationalist positions we have found questionable. (SSTS 42f)

Now, what would it be for us to treat the claim that truth and rationality and knowledge are relative to cultures as a hypothesis? What would count as evidence for its truth, and in what sense of the word 'truth'? What would count as evidence against it? Would anything be explained by assuming such an anomalous hypothesis?

Perhaps the last question is an unfair one. The logical oddity of a proposition is not necessarily a reason for excluding it from a scientific theory, if that theory has other virtues, such as predictive power, that would be lost by rejecting the logically odd proposition. (Suppose, for example, that its predictive power turns on a claim which, together with other claims of the theory, entails a contradiction.) But my other questions remain. What would it mean for the relativist's "thesis" to be reinterpreted as a "hypothesis?" In part it seems just to be the point that leads the relativist to retain her relativism despite the argument from self-refutation--*viz.*, there are compelling arguments against metaphysical realism, too, and the relativist sees subjectivity and objectivity as separated by an unbridgeable gap. Hesse distinguishes two "alternative situations":

> (i) We have some absolute criteria of knowledge in terms of which we can make absolute evaluations of belief systems including other parts of our own, and
> (ii) We have culturally relative criteria of knowledge in terms of which we can make relative evaluations of belief systems including other parts of our own. (SSTS 43)

---

[17] They might, of course, produce slightly different sorts of linguistic behaviour, but my claim is that the practical consequences of those differences will be negligible. As well, an astute fallibilist might well recognize occasions on which it is useful to imitate the speech of the relativist, and *vice versa*.

This is just the false dichotomy that makes relativism seem like the only alternative to metaphysical realism. There are no persuasive justifications of metaphysical realism; therefore, "it follows that there is no argument capable of establishing (i) rather than (ii)" (SSTS 44). We might as well, goes this reasoning, be relativists as realists, if this is the case, and since there are no deductive arguments to help us choose between the two, then why not see how each fares in light of the empirical evidence?

This brings the discussion back to my second question. What would count as evidence either for or against the relativist's hypothesis? Maybe the fact that there are widespread cultural differences of opinion about what is rational or true would confirm the hypothesis.[18] How would we "discover" that other cultures had different "criteria of knowledge" from ours--different standards of rationality and truth? In a sense we can discover this quite easily. Since different cultures value different ways of life, what is regarded as knowledge may well differ, and what is so regarded may well accord with that culture's reflective practices and procedures of justification and verification. But it is difficult to see why this should lead us to anything more surprising than the conclusion that different cultures take different things for granted and hold different things relevant. The relativist wants something more, but how can she get it?

The answer is that she can get what she wants quite easily--indeed, *far too easily*, because whether rationality and truth are relative to cultures is ultimately *for relativists to decide*. Whether the relativist's position is treated as a thesis or as an hypothesis, it remains the case that it is self-referential. So, if it is supposed, for the sake of "empirical" investigation of beliefs and the sources of their credibility, that truth and rationality are relative to cultures, then whether another culture has a different opinion of what is true or rational, and whether another culture makes use of necessary truths in its justification procedures, is itself true only relative to the culture undertaking the investigation. "Other cultures become," as Putnam notes, "logical constructions out of the procedures and practices" (PP3 239) of the investigating culture. It is "up to us" whether the hypothesis of the cultural relativism of truth and rationality is confirmed or disconfirmed by the evidence, since *what the evidence is* and what counts as (dis)confirmation is up to us.

We can make this point manifest by considering what happens if we shrink the culture to the size of a single individual. I am suggesting that there is no real difference between cultural relativism about truth and rationality and personal relativism--*i.e.*, subjectivism--about truth and rationality. Each displays the same logic, and each suffers from the same kind of incoherence. Putnam argues--and I think he is right--that the epistemic subjectivist is committed to treating each person as though she spoke a logically private language--that is to say, a language that is in principle incomprehensible to others. If I speak a logically private language, then you cannot learn it or understand it; I cannot teach it to you or impart an understanding of it. You might succeed in mimicking some, or all, of my utterances, but

---

[18] Better confirmation might be the discovery that "even if there are rational rules that are ultimately independent of social (and perhaps biological) causation" (SSTS 32), such rules are never appealed to in actual practices of justification. But I set this point aside.

as we saw in Chapter 1, §IV, the mere fact that we share a disposition to utter similar expressions on similar occasions is not sufficient for us to be following the same rule. (Indeed, it is not sufficient for us to be following any rule, because rule-following is typically a normative form of behaviour.) No matter how many points of coincidence there are in our uses of linguistic expressions, you do not *understand* my language, because it is open for me to reject your opinion of what counts as "going on in the same way." "That may be 'the same' for you," I might say, "but it's not 'the same' for me."

But, the subjectivist might respond, isn't this a *kind* of understanding?[19] Can't we say that each person has her own private understanding of what others mean when they speak? --However, this is compatible with my attributing any meanings (and beliefs and desires, etc.) whatsoever to another person and then claiming to "understand" what she has said. If we are to do any justice to the notion of understanding, then our account of it had better not entail the possibility that someone might understand another person's utterances and yet always be in practical conflict with that person over the meanings of her words and over the connections between her words and her behaviour. This is just what is not ruled out by a subjectivist account of understanding, for on this account understanding another's utterances amounts to interpreting them in whatever manner I want to. If truth reduces to what is "true-for-me," "true-for-you" or "true-for-some-arbitrary-person," then from any particular person's perspective there is no distinction between something's being true and her merely thinking it to be true, no distinction between her being right and her merely seeming to be right. What is "right" is up to her. "And," in Wittgenstein's words, "that only means that here we can't talk about 'right'" (PI §258). If what is true is relative to who I am, then only I am in a position to say what is "true-for-me." So, in principle, no one can correct my judgment, and *whatever* judgment I make will be "correct."

But all the same, it might be tempting to object, surely *I* can correct my own judgment, even if no one else can. I can change my mind from one time to the next, recognizing, *e.g.*, that an earlier judgment was mistaken, even though I thought at the time that is was true-for-me. So there really is a logically private fact of the matter about what is true-for-me, and there still remains a distinction between my being right (for me) and my merely thinking that I am right (for me). --But what are these logically private facts? The idea that I could speak a logically private language, a language which in principle "only I myself can [or, indeed, could] understand" (PI §256) is just the idea that the existence and nature of minds transcends the epistemic capacities of other minds, the idea that solipsism is a real possibility. *It is a form of metaphysical realism about other minds*!

Well, so what? Let's return to the idea that somehow there could be a logically private fact about the way I use 'true' and 'rational'. What I take to be the important point conveyed by Wittgenstein's argument against the possibility of a private language is that in the absence of there being some objective fact of the matter about the way in which I generally use words--"objective" in the non-metaphysical sense discussed in Chapter 2--there could not even *seem* to me to be such a fact. Why not? Because that there seems to me to be a fact of

---

[19] Thanks again to Bruce Hunter for this twist.

the matter about the use of my words--or yours--is not a mere sensation like a pain or an itch. There is no sense to the idea of what we might call "structureless seemings," or perhaps, "seamless seemings"--*i.e.*, "seemings" without some specifiable content. If it seems to me that there is some fact of the matter about how I use 'true' and 'rational', then this is not just an aura that accompanies my thoughts. Rather, there seems to me to be a fact of the matter *in some specifiable, describable, learnable way.* What such "seeming" depends on is my memory of having applied 'true' and 'rational' to particular cases in the past, and such cases are public phenomena.[20] Hence, it is open for anyone else to witness my applications of 'true' and 'rational' and so learn what I mean by them. But then, I am not speaking a private language, and truth and rationality are not relative to persons, because we can correct, or rationally dispute, each other's uses of 'true' and 'rational'.

If by contrast, the relativist tries to insist that there is no objective content to my "seeming" to use words in some particular way, then this seeming is just a brute-sensation. Imagine that every now and then when I make an assessment of "truth" or "rationality" I get an itch over my left eyebrow. Could this constitute a *reason* for withdrawing my judgment and saying, "I'm sorry--I was mistaken"? Should I say, "In the past when I have had such an itch I subsequently discovered that I had been mistaken about my concurrent judgment"? Should I then call this itch the sensation of "seeming to be wrong?" How would I know what it meant? How would I have subsequently discovered that I had been mistaken about my concurrent judgment, other than by resorting to some stateable standard of correctness? If this argument is compelling, and if there is no logical difference between subjectivism and cultural epistemic relativism, then relativism is a deeply incoherent position.

III. The Relativist as Realist: Cultural Essentialism

The relativist is inclined to resist the conclusion of the preceding section by seeking a fundamental difference between norms in the case of the individual and norms in the case of the group. If such a difference can be found, then epistemic relativism, it is thought, can avoid the incoherence of subjectivism. Thus, Anne Seller, whose views I shall examine more thoroughly in Chapter 5, wishes "to distinguish relativism, which," she thinks, "may be a coherent position, from subjectivism, which is certainly not" (RVR 170). Responding to criticisms of the Strong Thesis, Hesse objects to characterizing relativism as implying that "truth and validity [are] merely 'up to us'." "[I]f something is a function of a cultural situation," she writes, "it is certainly not wholly under our individual control, and therefore not just 'up to us'" (STSS 43). The distinction between the opinions of an individual and the standards of the culture to which she belongs is an important one for Hesse, but it is unclear in what the difference consists. Similarly, in response to the claim that

---

[20] Or at least some of them are, and while the Cartesian sceptic might want to press the case that all of my apparent public uses of language have really been private imaginings, this position has already been dealt with, and the subjectivist who claims not to be a sceptic will not want to challenge my claim of publicity.

shared basic needs and capacities might provide a basis for shared concepts and non-relative truth and rationality across cultures, Barnes and Bloor object that this reasoning depends on a "muddle between social and individual accomplishments" (RRSK 32). The sociology of knowledge examines "knowledge" as a "cultural product," a set of "collective," not "individual, representation[s]" (RRSK 32).

The great irony of these attempts is that it is precisely at these points that the relativist is most transfixed by the powerful myth of culture. The epistemic relativist, like the conceptual relativist, is a metaphysical realist about cultures--or, if you like, a cultural essentialist, and this position strengthens the parallel between the individual and the group. The cultural essentialist sees cultures as uniform and distinct; they are treated as "Cartesian Leviathans," composed of many individuals, but not suffering any important internal divisions, despite this fact. The many individuals are needed for the illusion that cultural relativism differs from personal relativism, but these many individuals' many differences from each other must ultimately be ignored, or the conceptual relativism on which epistemic relativism depends will be eroded into mere practical differences by the consequent fuzziness of the boundaries of cultures.

To clarify why this is so, we need to think for a moment about the relation between cultures and conceptual schemes. Perhaps the easiest way of expressing the relation between a conceptual scheme and a culture is to say that a conceptual scheme is part of a culture. If we share a culture, then we share a great many things, but among them will be a conceptual scheme, and in all likelihood, a language. I argued in Chapter 3 that there is no logical barrier to understanding across conceptual schemes, though there may well be significant practical barriers, which arise from the fact that practices of communication, linguistic or otherwise, are rooted in ways of life--and ways of life differ. Alternative conceptual schemes, to paraphrase Rorty's remark, are not like alternative geometries, which are *designed* to be incommensurable.[21] Whether or not holders of different conceptual schemes are able to understand each other is not a matter of logical form or of the intertranslatability of their languages or dialects, but of how capably they negotiate their differences. All this, I suggest, applies *mutatis mutandis* to different cultures.

The further consequences, of course, also apply. So, there are no clear boundaries between cultures, no necessary and sufficient conditions that will distinguish one culture from another culture. Similarly, there are differences within cultures. A conceptual scheme, to reiterate my earlier point, is not a formal, rigid structure, but a motley assortment of related practices, which differ in as many ways as they converge.[22] A culture, resting in part on its conceptual scheme and the practices and institutions in which that scheme is rooted, is similarly variegated, made up of various "sub-cultures" that differ no less than they are similar. Culture, as Ross Chambers

---

[21] See Rorty (1991), p.26.

[22] See Chapter 3, §V.

puts it, "is infinitely divisible and at the same time endlessly expandable."[23]

What makes these different sub-cultures parts of a broader culture is--in part--that their differences are not perceived by their members to be relevant. This is not the only criterion, and neither is it an indefeasible one. A culture that is subordinate to a dominant culture may well be perceived by members of both cultures to differ in ways that do matter. But to the extent that the character of the dominant culture is structured by its *dominance* over the subordinate culture--with the reciprocal relation holding for the character of the subordinate culture--we might well want to regard the two cultures as parts of a broader culture. My main point, however, is that there is no one similarity that can be found amongst all the sub-cultures of a culture. To adapt Wittgenstein's figure, the strength of the thread of a culture "does not reside in the fact that some one fibre runs through its whole length, but in the overlapping of many fibres" (PI §67). Cultures and sub-cultures bear a "family resemblance" to one another; they are situated in "a complicated network of similarities overlapping and criss-crossing" (PI §66). Intra-cultural differences are not different in kind from inter-cultural differences, though the latter may at first contact seem more dramatic because of their unfamiliarity. But from the moment of contact neither culture is quite what it was, because any attempt to compare "our" language with another, and so, "our" conceptual scheme with another, is an extension of "our" language and concepts which alters the very things being compared. Over time, the boundaries that geography and history allowed between cultures wear away, and slowly they may begin to amalgamate into a new culture. But the amalgamation need never be carried so far as to be judged a single culture (who is doing the judging is quite relevant), and even if it is, the differences that were thought to divide the two need not disappear, though they will be altered.

However, neither Hesse nor Barnes and Bloor treat culture in this way. Cultures seem to be, for them, internally homogeneous and essentially different from each other. Or, if two cultures are the same, they are essentially the same, and what may have appeared at first relative to one of them is relative to the two of them taken together. Thus, Hesse responds to Steven Lukes' variation on Davidson's argument against incommensurability as follows:

> [H]is arguments ... depend only on the unobjectionable view that we could not understand the other group's language or even know that they are asserting or arguing at all, unless they have some criteria of truth and validity in common with us. But even if this were true, it does not show that these criteria are in any sense external or 'absolute', only that they are relative to at least our pair of cultures, rather than to just our culture. (STSS 43)

The referent of 'this' is unclear. It could be either the claim that a particular culture shares "some criteria of truth and validity" with ours or the full conditional. Since the latter is "unobjectionable," I

---

[23] Chambers (1992), p.16. (I am grateful to Prof. Chambers for supplying me with a copy of his paper.) "*Indefinitely* divisible" might be more accurate than "infinitely divisible."

assume that Hesse has in mind the former. She is saying, then, that there might be other "groups" that had different criteria of truth and validity from our own and, as a consequence, whose language we could not in principle understand. This is just the old dualism of scheme and content that we have seen to be incoherent.

Hesse is more explicit about this possibility earlier in her essay when she offers the example of

> ... Martians who are making visible or audible gestures and who are evidently in reasonable control of themselves and their environment, and yet we persistently fail to get started on translation of their sign system because no consistent sense can be made of any hypotheses we make about the signs for yes and no. (STSS 37)[24]

As I indicated in Chapter 3, we might have extraordinary difficulty interpreting the behaviour of beings whose communicative and perceptual capacities were radically different from our own. But this is a severe practical difficulty, not a logical barrier. Curiously, Hesse seems to undermine her example (though it is meant primarily as a response to positions slightly more aprioristic than the one I am taking) on the following page where she observes that "we are able to *extend* our understanding and our language in unpredictable ways to give intelligibility" (STSS 38). This view is incompatible with conceptual relativism, as I understand it, though, as we saw in Chapter 3, Feyerabend accepts it while calling himself a relativist.[25]

The same conceptual relativism is the fragile crutch on which Barnes and Bloor lean their epistemic relativism. They see no problem in saying that the mere existence of alternative cultures is unnecessary for relativism, because their arguments would apply "even if there happened to be just one, homogeneous, international community" (RRSK 27n15). But were there only one community or not, it could never be utterly homogeneous and still be a community.[26] This may seem like an

---

[24] Hesse's focus on 'yes' and 'no' is not, it should be pointed out, an indication of her Quinean influence, though such influence is evident elsewhere in her essay. Rather, she is responding to the claim that "at least the concepts of negation, contradiction (the 'yes/no' distinction), and entailment must be present in order for us to know that another system of signs is a language at all" (STSS 37).

[25] See, *e.g.*, Feyerabend (1988), p.197 and FR 272. I think that Rorty is thus correct in saying that "[Feyerabend's] project seems misdescribed, by himself as well as by his critics, as 'relativism'" (Rorty, 1991, p.28).

[26] Barnes is more careful in a later paper, already cited, acknowledging the existence of sub-cultures (Barnes, 1987, p.20), but he persists in accepting relativism, largely because he confuses the "open-ended character of concept-application" (*ibid.*, p.32) with the "community-view" of rule-following, examined and criticized in Chapter 1, §IV. The latter he seems to take as a reason for holding that different extensions of the "same" concept are on a par rationally, because the concept is indeterminate anyway, given its application only in a finite number of instances. What this would show, if it were

unfair over-interpretation of a philosophically innocent remark, and were it the only indication of Barnes and Bloor's views, my complaint would be tendentious. But it becomes another piece of the interpretive puzzle when we set it next to the same juxtaposition of views present in Hesse's essay: epistemic relativism and a belief that understanding another culture depends not on being able to translate its language, but on being able to "proceed in the way that native speakers do" (RRSK 37). If one subscribes to cultural relativism about truth and rationality, while also holding that understanding a culture is both possible and not merely a matter of translation, then I submit that this can only result from being in the grip of cultural essentialism. On Barnes and Bloor's view, and on Hesse's too, perhaps, when two cultures come into contact, their "essences" remain unchanged. For unless there were some logical or metaphysical barrier, what could prevent favourably disposed individuals of different cultures from trying to resolve or accommodate their differences through argument, criticism and a commitment to achieving a shared understanding? Unless we are held by the spell of metaphysical realism about cultures, then why should we be persuaded to treat our differences as anything but differences of opinion and practice?

Neither cultures nor their conceptual schemes are uniform and strictly bounded. So, the epistemic relativist cannot fall back on the greater *prima facie* plausibility of conceptual relativism to save her position. Nor can the epistemic relativist block the move from culturally relative rationality and truth to personally relative rationality and truth. The former is as incoherent as the latter, but the incoherence of the latter is manifest. And here we see that the relativist, whether she places her faith in conceptual relativism or in epistemic relativism, suffers, as the metaphysical realist does, from a kind of epistemic neurosis. For like the realist, if she is to express her view, she must base it on an assumption that ultimately undermines that view. Like the realist, she does not believe what she says, does not doubt when she says she is not sure. If she really did treat truth, rationality and concepts as utterly relative, then she could not seriously consider presenting arguments for her view.

This last point bears emphasis, since at least some relativists *do* take the time to argue for their positions. But what can such argument come to? If epistemic relativism is "correct," then what is true-for-the-relativist need not be true-for-me, and *vice versa*. If relativism is not true-for-me (the implication that the relativist should--but need not--draw from my criticisms of her position), then how can the relativist go about convincing me of her position? First, it would be unclear what she were trying to do. If she wants to avoid quick self-refutation, then she must avoid trying to convince me that relativism is true *sans phrase*. Maybe she should try to show me that relativism is true-for-her. But what grounds can she give me for believing this, unless all she means is that she believes in the thesis of relativism? (And then, I will repeat the diagnosis of epistemic neurosis.)

If I am to be convinced, then grounds or evidence must be supplied which do not count merely as reasons-for-the-relativist. Or--less

---

[26] successful, is that a relativist need not be a cultural essentialist, because she can acknowledge sub-cultures by insisting on the indeterminacy of meaning. For reasons I have already sketched, however, this hardly seems an improvement.

smugly--it must be conceded that there could be such grounds. How can this be accomplished without agreeing to give non-relative reasons? *Prima facie* there are two possible ways. The relativist might try to convince me that relativism was *true-for-me* by giving an immanent critique of my own position. But this just reduces to another description of the ongoing debate between relativists and non-relativists, and I have already indicated why the stage-sets of relativism collapse when it performs this play.

Secondly, the relativist might revert to the claim, considered above, that relativism is best thought of as an "attitude." To lend weight to this claim she might invite me to see the world through her eyes in the hope that, if I try to live the life of a relativist, as it were, then I will come to understand relativism and perhaps embrace it. But understanding relativism is not like understanding another culture; it is not clear that any particular way of life, or any collection of ways of life, attaches strictly to the relativist's thesis. Maybe this is not quite right. Conservative critics of the left and of feminism sometimes argue that a culture or sub-culture of relativism really is manifested by political radicals, and it is clear that one important set of motives for relativism is political or ethical in nature. However, one can sympathize with a variety of radical political concerns and be an active participant in this sub-culture without being at all sympathetic to relativism. One can be a conscientious fallibilist who accepts a coherent form of realism--such as I have been trying to present. In the next chapter I shall argue that the concerns of many politically motivated relativists (but not all, there are *conservative* relativists) can be met by abandoning metaphysical realism and scientism, without embracing relativism.

Chapter 5: Relativism, Realism and Scientism

In the preceding chapter I explored some standard arguments for epistemic relativism, suggesting that to the extent that they are plausible they constitute arguments against metaphysical realism, but not for epistemic relativism. As we saw, epistemic relativism tends implicitly to rely on the greater initial plausibility of conceptual relativism, resorting to an application of the metaphysical realist's notion of objectivity to cultures and conceptual schemes. I shall now take up another kind of argument for relativism, which I shall call the "ethical-political" argument. Fearing that metaphysical realism devalues cultural and other differences, this sort of relativist looks upon her position as a way of celebrating those differences and as a way of responding to ethnocentrism. Despite the illogic of this position, it stems from concerns that deserve to be taken seriously. I shall argue that adopting epistemic relativism in fact exacerbates those concerns, and that the politically motivated relativist tends to conflate the parts played by realism and scientism in motivating the attitudes to which she objects. It is interesting to ask, however, why the relativist conflates these two philosophical positions, and I shall explore the idea that realism and scientism might be linked conceptually. I shall argue that in fact they are not, but that the relativist's conflation might be explained by one or both of two historical alliances of realism and scientism: their frequent concurrence in contemporary analytic philosophy and their long-standing association with each other in varieties of Marxism. These claims I shall present as tentative hypotheses in need of further study before they can be taken as well-confirmed. However, during the course of my discussion I shall suggest that my position is not vulnerable to scientism in the way that metaphysical realism might seem, and that a non-scientistic form of *Ideologiekritik* is compatible with the views I have advanced in this and earlier chapters. Thus, at the same time as I offer a brief account of the critical theorist's journey to relativism, I hope to tempt the would-be relativist with glimpses of the not so-distant shores of internal realism.

I. The Ethical-Political Argument

I indicated in Chapter 4 that relativism is sometimes embraced, not so much for positive conceptual reasons, but because it is seen as the only alternative to metaphysical realism. It is also regarded as just as intellectually respectable as realism, because realism is held to be incoherent for reasons not unlike those I have advocated in earlier chapters. By ordinary standards of "intellectual respectability" this is not to say much, but if one is of the conviction that epistemic relativism and metaphysical realism are the only choices, then such matte-reasoning acquires a certain dull lustre. One factor that can tip the balance in favour of relativism in the mind of a dichotomizing thinker is the perceived political or ethical perniciousness of realism. Such perceptions are not universal, but they do exercise a certain hold over some thinkers who believe that philosophy ought to be put to work in the service of struggles against power and injustice--or conversely, who believe that philosophy ought simply to

avoid taking sides, either with the radical critic, or with the "traditional theorist"[1] whom the radical critic opposes.[2]

The claim that relativism is morally or politically superior to metaphysical realism I shall call the "ethical-political" argument. It is an argument that has been recognized in some form by a number of critics of relativism. Thus, Charles Taylor writes that relativism is "tempting," because "It takes the heat off; we no longer have to judge whose way of life is superior."[3] And Putnam observes that the relativist wants "to convince us to stop destroying primitive cultures by attacking our belief in the superior rationality and morality of our own" (RTH 161). Despite good intentions, Putnam goes on to say, the relativist "has chosen the wrong argument" (RTH 162). Indeed, the argument that can easily be bent to imperialistic purposes, since if we have only our own "truths" and "rationality" to consider, then there can be no arguing against us, if what we hold rational includes forcing our hegemony on others. If we assume epistemic relativism, "why should we not destroy whatever cultures we please" (RTH 162)?

I am inclined to agree that politically motivated relativists have "chosen the wrong argument." But while critics have correctly identified this motivation for relativism and the problems that it faces, there is little sustained discussion of why the ethical-political argument might seem attractive. What are the assumptions that lead a radical theorist to relativism, and can the concerns of the politically motivated relativist be met without resorting to relativism? Closer scrutiny, I believe, reveals a thinking that not only dichotomizes the subjective and the objective, but suffers from a lack of clarity concerning the relationship between realism and scientism. The worries of such relativists can to a large degree, perhaps entirely, be met without embracing relativism.

Whether one thinks that this discursive gap is worth filling may depend in part on whether one feels the political attraction of relativism. To some the politics of the radical theorist will just seem controversial. But while I sympathize with some of the claims that politically motivated relativists make, I shall not be concerned

---

[1] The term is Horkheimer's. See Horkheimer (1982), pp.188-243.

[2] Barnes and Bloor think that relativism clears the intellectual path of political biases, making way for disinterested research:

> A plausible hypothesis is that relativism is disliked because so many academics see it as a dampener on their moralizing. A dualist idiom, with its demarcations, contrasts, rankings and evaluations is easily adapted to the tasks of political propaganda or self-congratulatory polemic. *This* is the enterprise that relativists threaten, not science. (RRSK 47n44)

How "disinterested research" (RRSK 47n44) is possible, assuming that one's own culturally relative standards of rationality are the fixed glass through which the world is viewed, is somewhat mysterious. Barnes has more recently said that interests may be "constitutively involved in the processes of concept application by which the knowledge of a community persists and grows" (Barnes, 1987, p.29f).

[3] Taylor (1982), p.99.

to argue for any of those claims here. Suffice it to say that I think that it is worth trying to understand why relativism should seem so tempting for ethical-political reasons, and I take myself to be making suggestions to the relativist that would better serve her political concerns at the same time as I clarify for the non-relativist the attraction that relativism holds for those same radical theorists.

Some philosophers wear on their sleeves their political commitment to relativism. Anne Seller criticizes realism under the heading of "rational-scientific epistemology" for what she perceives to be its anti-democratic stance:

> [A]lthough this epistemology, which I call rational-scientific, is politically appealing (it enables us to say to the sexist 'you are wrong') it also raises political problems. First, it is an élitist epistemology. Only some women have the resources (time, library, etc.) to conduct such research, other women will simply have to accept it on authority. ... Secondly, women have often experienced the scientific-rational approach as oppressive both in its process and in its findings. (RVR 170f)[4]

Related worries with respect to cultural imperialism are cited by Paul Feyerabend as a source of his disquiet with realism and his advocacy of "relativism."[5] Feyerabend embraces relativism as "a weapon against intellectual tyranny" and "a means of debunking science" (FR 19), or at least its title to paradigm of knowledge.

An examination of the concerns expressed in versions of the ethical-political argument reveals a failure to distinguish metaphysical realism from the doctrine that all and only science is to count as knowledge. The latter view is *scientism*, and it often assumes, not only that science is the model of rationality, but that the physical

---

[4] Seller's criticisms are directed against a variety of realists including other feminist philosophers such as Jean Grimshaw. For Grimshaw's position see Grimshaw (1986), esp. Ch.3.

[5] As I noted earlier, I am not convinced that Feyerabend really is a relativist, at least not in the sense of the word I have been using here. Nonetheless, this remark is enlightening to precisely the extent that a number of relativists, including those whose work I have been discussing, seem unclear about the consequences of their views and about their views themselves.

sciences should be the model for the other sciences.[6] Seller is quite candid about conflating realism and scientism. Realism, she says,

> is the view that there is an objective order in human affairs, independent of people's beliefs about it, which can be discovered by some methodology generally characterised as rational and scientific. Thus, on this view, both Marxists and positivists might be characterised as realists because they believe in a social reality discoverable by the use of a method they specify as scientific. (RVR 183n2)

These remarks present an array of difficulties. According to Seller, realism is "a view that ... is best understood in opposition to relativism" (RVR 183n2). But her account of "realism" does not distinguish metaphysical realism from internal realism, because it mentions only independence from people's actual beliefs and not independence from people's epistemic capacities. This suggests a dichotomous separation of realism and relativism. (Notice, however, that it is "social reality" and "human affairs" that concern Seller, not "the external world" in general.)

But more important here is her qualification of this independent reality as knowable "by some methodology generally characterized as rational and scientific." The adoption of a "scientific method" is not logically decreed by the metaphysical realist's notion of objectivity,

---

[6] I have some sympathy for Feyerabend's claim that "science," properly understood, is really a "fictitious unit" (FR 36).

> Scientists have taken ideas from many different fields, their views have often clashed with commonsense and established doctrines, and they have always adapted their procedures to the task at hand. There is no one 'scientific method', but there is a great deal of opportunism; anything goes--anything, that is, that is liable to advance knowledge as understood by a particular researcher or research tradition. (FR 36)

Even within the natural sciences there is considerable variation of appropriate procedures, and where science ends and some other category of knowledge begins cannot be sharply delineated. Complaints with scientism should, thus, be aimed not at the rationality of the natural sciences, but at the elevation of an over-simplification of this model to a standard against which all knowledge-claims should be assessed. I think I detect a similar point in Campbell (1993).

My endorsement of Feyerabend's point might seem to undermine my determination to distinguish, *e.g.*, the normative from the empirical. However, to say that there is no sharp boundary to be drawn between science and non-science or between the normative and the empirical is not say that there is no boundary at all. As well, the criticism neglects the very point I have just raised--*viz.*, that scientism is not merely the elevation of science to the level of overriding standard of rationality, but the elevation of an oversimplified, "fictitious" view of science to that level.

or by the weaker conception of objectivity argued for in Chapter 2.[7] In fact, on inspection, Seller's arguments, ostensibly for relativism, prove to be for the most part, arguments against scientism. Thus, noting the weight that some feminists place on exposing "scientific" claims such as "a woman's womb withers if she uses her brain" (RVR 170),[8] Seller worries that debunking these claims makes it tempting to adopt "an epistemology based on that scientific approach" (RVR 170), as though recognizing the usefulness of science in specific cases implied ruling out other kinds of knowledge-claims. It is this hasty generalization, I think, that leads her to the objections encountered above: that "rational-scientific epistemology" is "élitist" and that "women have often experienced the scientific-rational approach as oppressive both in its process and in its findings" (RVR 171).

I do not dispute the second of these claims, but I think it is worth pondering the idea that the role of science in women's oppression depends on (at least) two factors: first, the absence in the past of scientific feminist critiques, and secondly, the privileging of science as an authoritative discourse--*i.e.*, scientism. Eliminating the latter factor also eliminates the worry that scientific feminist critique entails a monolithic "rational-scientific epistemology." Only if one supposes that "science" must either be the paradigm for all knowledge or be abandoned for some other general paradigm, does it make sense to worry generally about what to do with critiques that draw on existing scientific paradigms.[9] And only if one supposes that science (or another technical discourse) must take centre-stage or not appear at all, need one worry about having "[o]ne epistemology for the élite, another for the masses" (RVR 183n2). Abandoning the view that all and only science is knowledge (where science is often understood according to the model of physics and chemistry) undermines the source of Seller's charge of élitism. If science is élitist, then that is largely because there is a tendency in European culture and its descendants to see a simplistic image of science as the model of rationality. The barriers that women and some visible minorities face to entering scientific professions are a related factor, but the relation is dialectical, and such barriers are not exclusive to the sciences. The élitism of science consists partly in its exclusion of most women and many others, but that exclusion is also perpetuated because science is, in certain respects, élitist, as are other disciplines. Overcoming the élitist split that worries Seller entails neither having all women become scientists, with the residual worry that many "women have often experienced the scientific-rational approach as oppressive," nor abandoning science and scientific critiques of particular cases of sexist thinking.

A further worry that Seller raises for "realism" displays this same conflation of scientism with realism and their further conflation with the Cartesian quest for certainty. In contrast to traditional epistemology, she writes, "We do not have necessary and sufficient

---

[7] I shall return to these claims below.

[8] See Grimshaw (1986), p.101. On this and other bizarre proposals see Ehrenreich and English (1978), pp.125-131 and *passim*.

[9] A critique may be flawed, of course, but there is no necessary conflict between scientific method and feminism. See Campbell (1993).

tests of the truth, which we can individually apply, such as Descartes and so many since him have sought, but a process of conversation which may allow the truth to emerge ..." (RVR 179). This complaint needs explication. Her main concern, I think, is to distinguish the method of feminist "consciousness-raising" from the methods that a variety of epistemologies might recommend on the basis of treating epistemology as centred on an independent, individual knower. But rolled together here are claims about the unity of method, the reliability of the results of method, and the nature of feminist method.

That "consciousness-raising" is the *method* of feminist theory is a claim closely associated with Catharine MacKinnon. "Feminism," she writes, "does not appropriate an existing method--such as scientific method--and apply it to a different sphere of society to reveal its pre-existing political aspect" (FMMS 535). Rather, she claims, feminism's distinctiveness lies in its emphasis on dialogue and the discussion of women's experiences, discussion that shines new light on those experiences. MacKinnon's case for viewing consciousness-raising as a *method* rests on a list of epistemic advances that she takes feminism to have made, largely through shared discussion. Seller offers a similar, but briefer, account:[10]

> Women's oppression has partly been understood in terms of the silencing of women, the denial of their experience as valid, or the treatment of it, when discovered, as neurotic. The woman who failed to find satisfaction in the fulfilment of domestic duties or who did not want to have babies was treated as a suitable case for treatment. The apprehension that such women were not sick but oppressed by a false view of what they should be came about only through women sharing these feelings and experiences with each other. (RVR 176)

That consciousness-raising could be a method for acquiring knowledge is intimated by the observation that two heads are better than one. Informal empirical evidence suggests that one does learn things by talking to others. Indeed, one can learn things about oneself by talking to others. In part, this is because the *differences* in another's experiences allow her to see my experience differently from the way I do, and so, to give me new descriptions of those experiences and of myself.[11] And to the degree that we are similar, her sharing certain views with me may reinforce my conviction about matters regarding which I had only the most tentative and insecure suspicions,

---

[10] Seller might well demur from MacKinnon's assessment of consciousness-raising as a "method," perhaps on the ground that "method" is too closely associated with "science." This, however, is speculation on my part. Seller also differs from MacKinnon insofar as the latter does not embrace relativism--though some critics might accuse her of slipping into it.

[11] In multiplying heads, we do not simply multiply intelligences and confirmations, but we produce knowledge in the exchange of views, multiple, slightly different, sometimes opposed, and in the questioning, perhaps precisely because of our differences. (RVR 180)

or which I had even managed to hide from myself. Intersubjectivity is a mark of objectivity, though not a sufficient condition.

Part of Seller's complaint, then, is that doctrines of the unity of method, *e.g.*, logical positivism's doctrine of the unity of science, do not account for the epistemic importance of dialogue--at least insofar as they focus on the idea of a formulaic method for the natural sciences. A related complaint seems to underlie her use of the expression "necessary and sufficient tests for truth." However, this phrase introduces further complications. A *test* that is necessary and sufficient for truth seems both to give a necessary and sufficient *condition* for truth and to be part of a method for acquiring knowledge or justifying knowledge-claims. But a test that fits both of these descriptions is part of an *infallible* method, one which if followed with adequate care, leads to *certain* conclusions. This might explain Seller's mention of Descartes, but it also makes it hard to discern whether she is unhappy with claims of certainty, with the idea of a method that offers necessary and sufficient conditions for *justifying a belief*, or with the idea that there is some set of conditions necessary and sufficient for a belief or proposition's being true.

Consider first the worry about certainty. This is not a worry about scientism, but neither is it a worry about realism. We saw in Chapter 4 that the relativist might present her position as expressing a kind of attitude, a reluctance to make or to accept pronouncements that masquerade as authoritative and beyond criticism. But I also suggested that this attitude is difficult to distinguish from a conscientious fallibilism, a recognition that I can be wrong about any particular claim that I make and a consequent modesty about the reliability of my beliefs. This attitude is quite compatible with a coherent realism--such as, I claim, my own. Indeed, nothing about *metaphysical* realism requires the acceptance of a method whose results are certain in the way Descartes hoped, immune to error--nothing, that is, beyond the epistemic neurosis from which it suffers and which it shares with relativism. Some metaphysical realists, as we have seen,[12] try to rescue their positions from the sceptic by denying that knowledge requires certainty. To demand certainty, they say, is to set the standards of knowledge too high. So, the rejection of Cartesian certainty as an epistemic standard is no argument for relativism.[13]

Now, it seems implausible to suppose that Seller has a complaint with the third of the options that might seem to be present in her remark about necessary and sufficient tests for truth. There is an utterly trivial sense in which even the relativist would probably want to acknowledge that some set of conditions is necessary and sufficient for truth--*viz.*, "*we have said what is true iff what we were talking about is as we have said it to be.*"[14] The relativist will want to relativize this statement, but not likely to abandon it.

Quite apart from the oddity of such a position, I think that the interpretive evidence simply indicates that Seller is concerned with

---

[12] Chapter 2, §IV.

[13] All this is compatible with claiming that the desirability of Cartesian certainty is difficult to make sense of without metaphysical realism. See Chapter 6, §I.

[14] Alston (1978-79), p.780.

something else. What I think she actually has in mind--in addition to the worry about certainty--is the rejection of the idea that there is some set of conditions, necessary and sufficient to *justify* a belief in any justifiable proposition. It is, as I suggested, a worry about the unity of method. Seller wants to deny, first, that a belief is justified if and only if it is arrived at, or subsequently confirmed by, *the* right method, where the right method prescribes a series of tests that any rational individual can, in principle, apply, and secondly, that feminist consciousness-raising relies on such a method. The latter denial we examined briefly above. But, as I indicated, neither it nor the denial of the claim that a belief is justified if and only if it is arrived at, or subsequently confirmed by, *the* right method constitutes an argument against realism or for relativism. It is, thus, doubtful that Seller's worries can be alleviated only by relativism. Indeed, it is not even clear that one need abandon metaphysical realism in order to meet these concerns, although I have offered other arguments for taking that step.

Is Seller's criticism of realism based simply on a confusion, then? One point that she makes does seem relevant to realism, rather than to scientism. Discussing epistemic authority, she argues that the notion of "cause" is interest-relative: "we use value judgments in picking out certain conditions as precipitating events" (RVR 174). Does this point require epistemic relativism?

I tend to agree with Seller's claim that "cause" is a notion relative to our interests. She illustrates the point with an example in which she claims that cases of childhood-leukaemia are caused by a nuclear generating station, while nuclear experts respond that the cause in question is a lack of proper nutrition among the afflicted children (RVR 174). Both claims may be "true," she says, because each condition may be insufficient in the other's absence for an increased incidence of childhood leukaemia. What counts as "the cause" here depends on who should assume responsibility for the problem. This interest-relativity of causes, she seems to suggest, is a reason for embracing epistemic relativism.

Seller's position is similar to Collingwood's claim that a cause is

> *... an event or state of things which it is in our power to produce or prevent, and by producing or preventing which we can produce or prevent that whose cause it is said to be.*[15]

"[I]f my car fails to climb a steep hill," he writes, "I shall not consider my problem solved by a passer-by who tells me ...that ... more power is needed to take a car uphill than to take her along the level."[16] What the passer-by says is correct, but it does not answer the driver's problem, because the driver cannot do anything about the steepness of the hill, whereas he might be able to do something with his car. If the driver "had been a person who could flatten out hills

---

[15] Collingwood (1940), p.296f. Collingwood discusses several other senses of 'cause', too. Thanks to Bruce Hunter.

[16] *Ibid.*, p.302.

by stamping on them,"[17] then the advice would have been appropriate, but, as is, it constitutes no more than a bit of levity, at best.

The relativity of causes is fairly strict on Collingwood's account, for it is *only* what is at present actually within my control that he considers to be a cause for me. "[O]nly a person who is concerned with producing or preventing a certain kind of event can form an opinion about its cause. For a mere spectator," he says, "there are no causes."[18] However, we do seem able to understand why the driver's question is not answered by the rib-tickling passer-by, and we are also able to understand Seller's question about the cause of leukaemia. Yet, in neither case need we be in the situation described in order to understand the problems involved.

What this shows is that the interest-relativity of causes is no argument for epistemic relativism. Only if we could not, in principle, understand the situation of the driver whose car cannot make it up the hill, would we be justified in supposing that something were true for him that were not true for us. I have already argued against such principled barriers to understanding. The interest-relativity of causes does not require relativism, though it may cast doubt on metaphysical realism, or at least the correspondence-theory of truth, whose contemporary proponents want to reduce reference to a set of objective--*i.e.*, non-relative, non-intentional causal relations.[19]

II. Realism and Scientism: Necessary Connections?

In the preceding section I made two major points about Seller's reasons for embracing relativism. First, many of her concerns have more to do with scientism than with realism--metaphysical or otherwise. And secondly, her points that do not mistake scientism for realism do not dictate epistemic relativism. But could there be a subtle connection between metaphysical realism (or, indeed, my own position) and scientism that I am missing?

Seller is not alone in linking realism and scientism. Catharine MacKinnon writes of the metaphysical realist's notion of objectivity as though it were intrinsically linked with scientism.[20] According to MacKinnon, the inappropriateness of applying "science" to an analysis of women's oppression stems from the fact that science must draw a clear distinction between epistemic agents and a mind-independent reality, though the nature of that independence is left unclear:

---

[17] *Ibid.*, p.303.

[18] *Ibid.*, p.307.

[19] As noted in Chapter 1, §VI, Putnam invokes the interest-relativity of causation in just this context. See Putnam (1987), pp.37-40 and Putnam (1992a), pp.44-55.

[20] As I have noted I do not think that MacKinnon's position amounts to a form of relativism, though this is a contentious point of interpretation. I read many of her claims as hyperbolic, reminiscent of the occasionally reckless, but ultimately comprehensible, style (at least in translation) of such (varied) French thinkers as Sartre, Althusser and Derrida. She herself cites the influence of such French feminists as Hélène Cixous, Luce Irigaray and Julia Kristeva.

> The problem with using scientific method to understand women's situation is that it is precisely unclear and crucial what is thought and what is thing, so that the separation itself becomes problematic. (FMMS 527n23)

MacKinnon hopes to make a theoretical contribution to overcoming women's oppression by dissolving the objective-subjective distinction:

> Disaffected from objectivity, having been its prey, but excluded from its world through relegation to subjective inwardness, women's interest lies in overthrowing the distinction itself. (FMMS 536)

Elsewhere, she focuses her criticisms more directly on metaphysical realism, though the passages cited above make it difficult to say whether she intends to draw a distinction between objectivity *qua* independence from people's beliefs and objectivity *qua* independence from people's epistemic capacities:

> Cartesian doubt ... comes from the luxury of a position of power that entails the possibility of making the world as one thinks or wants it to be.[21]

Her view seems to go something like this: a position of relative power makes it possible for men to confuse "what [they] think" with "the way the world is,"[22] because while the world is taken as an independently existing thing that places constraints on our behaviour, men's behaviour towards women is in important ways *un*constrained. According to MacKinnon, it thus becomes possible for men to doubt whether women possess any independent reality. Just as the Cartesian sceptic fears the world to be an illusion--perpetrated by the mind or some other agency--so women may end up being viewed as male creations.[23] (Women, by contrast, may find the realities of oppression harder to doubt, though this will be a matter of degree.) My central point here is that

---

[21] MacKinnon (1987), p.58.

[22] *Ibid.*

[23] As Bruce Hunter has pointed out to me, this remark suggests a slippage from sceptical doubt to idealism. But this move is recognized in Kant's claim that the "transcendental realist" is driven toward idealism in order to avoid scepticism. See A 369. In this context MacKinnon draws an amusing parallel between Cartesian doubt about the external world and male doubt about female orgasms. Her point, I take it, is that in the latter case, as in the former, there is a genuine concern that no *reality* underlies the *appearance*. See MacKinnon (1987), p.58. It would be interesting to pursue the question of whether the epistemic neuroses of metaphysical realism have genuine neurotic correlates in the ways in which the powerful conceptualize the world and their subordinates. Is ideology a neurotic response to the fear of loss of power?

MacKinnon draws a link between the "male,"[24] "scientific" conception of *objectivity* and the increased chance of women's *objectification.*

In a different vein, but also assuming a link between realism and scientism, Feyerabend portrays relativism as a philosophical and political response to "the idea of Reason and the idea of Objectivity" which, he says, "have often been used to make Western expansion intellectually respectable" (FR 5). Feyerabend is by no means hostile to science, but he insists that "science" understood as the paradigm of rationality on the basis of its exact and authoritative methods is a "fictitious unit" (FR 36). The idea that there is *a* scientific method of the sort that worries Seller is an idea that Feyerabend sees as distorting science, imposed on it by a reverence for Objectivity and Reason. Objectivity is taken by Feyerabend to have pernicious results, and he clearly treats it as logically distinct from science:

> To say that a procedure or a point of view is objective(ly true) is to claim that it is valid irrespective of human expectations, ideas, attitudes, wishes. This is one of the fundamental claims which today's scientists and intellectuals make about their work. *The idea of objectivity*, however, is older than science and independent of it. It arose whenever a nation or a tribe or a civilization identified its ways of life with the laws of the (physical and moral) universe and it became apparent when different cultures with different objective views confronted each other. (FR 5)

Here is yet another instance--possibly deliberate--in which the two senses of objectivity that I have tried to distinguish are run together. Whatever the status of this conflation, however, it is plain that Feyerabend would regard metaphysical realism as objectionable quite apart from any relation to science and scientism, simply in virtue of its attachment to Objectivity. Just as the standpoint of objectivity, in MacKinnon's story, is supposed to make it easy for men to confuse what they think with the way the world is, so Objectivity, in Feyerabend's story, is supposed to allow a civilization to confuse "its way of life with the laws of the ... universe."

But what are Feyerabend's grounds for blaming objectivity? He lists three responses to the idea of objectivity. With two of them, "opportunism" and "relativism" he expresses some sympathy. The former, he writes "is closely connected with relativism; it admits that an alien culture may have things worth assimilating, takes what it can use and leaves the rest untouched" (FR 86). It is only the remaining

---

[24] The idea that there is a certain "maleness" to concepts such as "objectivity" and "reason" should not be taken as an assertion of a strict dichotomy between men's thinking and behaviour and women's thinking and behaviour, based either in strict biological determinism or strict cultural determinism. (MacKinnon is not always clear about this, especially with regard to the latter determinism, but that may have something to do with her primarily intended readership.) Good discussions of the "maleness" of philosophy, which avoid slipping into attributions of essential differences include Grimshaw (1986), pp.36-74 and Lloyd (1984). However, among feminists there is a debate about whether even such limited categorizations of philosophy, reason, etc. as "male" are justified. Thanks to Edrie Sobstyl for this last point.

response, "persistence," that incurs Feyerabend's displeasure, and even here his disapproval (which I infer) is not unequivocal:

> One reaction was *persistence*: our ways are right and we are not going to change them. Peaceful cultures tried to avoid change by avoiding contact. The pygmies, for example, or the Mindoro of the Philippines did not fight Western intruders, they did not submit to them either, they simply moved out of their sphere of influence. More belligerent nations used war and murder to eradicate what did not fit their vision of the Good. (FR 5f)

But persistence in its belligerent guise seems not all that different from Feyerabend's account of Reason: "This belief ... may be formulated by saying that there exists a right way of living and that the world must be made to accept it" (FR 11). It would seem, then, to be the *combination* of some form of realism with some authoritative (or authoritarian) form of "Reason" that troubles Feyerabend, and realism enters the picture, only because it facilitates hegemonic Reason. At this point, Feyerabend's views strongly resemble MacKinnon's, for part of her concern is that objectivity facilitates *objectification*, and in her opinion science is a key method of objectification: "What is objectively known corresponds to the world and can be verified by pointing to it (as science does) because the world itself is controlled from the same point of view" (FMMS 538).

We have been swept along by a torrent of quotations, and it would be appropriate to step ashore for a moment and try to make some sense of the passing stream. I began with this worry: is realism (metaphysical or otherwise) the real object of the politically motivated criticisms that lead some thinkers to adopt relativism? The answer that is beginning to emerge from the flood is a qualified "no"--"no" because the primary focus of the criticisms I have examined still appears to be scientism; a *qualified* "no" because these criticisms seem also directed at forms of realism that are regarded as preconditions of scientism. But is any form of realism necessary for scientism? Is any form sufficient? I asserted earlier that this is not the case, but perhaps closer consideration is required.

The latter point is easier to settle. The metaphysical realist, recall, holds that the world's existence and nature are independent of people's capacities to know or describe them. But holding such a view is compatible with believing in similarly objective grounds of moral or aesthetic value and with thinking that science tells us nothing of morals or aesthetics. One could imagine this to be a variation on Platonism, itself a form of metaphysical realism,[25] or we might think

---

[25] See Plato (1974). Scholars of Plato might well take issue with this claim, but I have in mind a commonly received view: according to Plato, the Forms are the only reality, while the world of sense is nothing but illusion. The Forms are held to exist independently of the world of sense, depending only on the Form of the Good, which in turn is self-sufficient. Plato does think that we have innate knowledge of the Forms, which we come to "recall," first by being reminded of the Forms by aspects of the world of sense and then, if lucky and suited, through education in dialectic. However, this innate knowledge is a contingent fact about us, not a condition on which the existence and

of the ethical and aesthetic non-naturalism of G. E. Moore and the Bloomsbury group. If we contemplate an "exceedingly beautiful" world and "the ugliest world [we] can possibly conceive," says Moore, we should conclude that it is better that the former, not the latter, exist even "supposing them quite apart from any possible contemplation by human beings."[26] Neither of these views is scientistic, but each is committed to a very strong conception of objectivity. Metaphysical realism is also the implicit metaphysics of strains of religious fundamentalism, whose supporters seldom feel much sympathy for the claim that all and only scientific knowledge counts as knowledge.

However, these are examples of élitist epistemologies--or are likely to appear so, at least[27]--even if they are not scientistic. Dialectic is a discipline for which, according to Plato, few are suited, and his *Republic* is premised on the idea that such élitism is the best way of choosing a ruler. Fundamentalism, similarly, often takes some holy text as the final authority and only the chosen few have access to knowledge by way of the word of God, or of Allah, or of whoever.

These facts invite interpretation in light of Feyerabend's claims that "objectivity ... is older than science and independent of it" and that Reason is the "belief ... that there exists a right way of living and that the world must be made to accept it." But even if we accept Feyerabend's account of Reason and Objectivity, it still does not follow that objectivity--*i.e.*, realism--is at fault here. Feyerabend's Reason is one response to the discovery of objectivity, according to him, and to say that relativism and its companion, opportunism, exhaust the field of responses is to neglect the possibility of a non-authoritarian pluralism that is not committed to relativism. Indeed, I think that Feyerabend's own position is better characterized as such.

We must consider, then, whether realism is necessary for scientism, and I do not think that a positive conclusion is forthcoming here either. It seems clear that some variety of positivism would easily fit the bill for a philosophical position both scientistic and non-realist. The logical positivists rejected the account of objectivity embraced by the metaphysical realist, offering in its place an account similar in respects to the one that I have endorsed.[28]

Such a conception was evident in my brief examination of Carnap's views on "frameworks" in Chapter 3. The concerns of the Cartesian sceptic can get no hearing from Carnap, for such doubts about the external world question the linguistic framework of "thing-language," which is presupposed by the very expressions used by the sceptic to formulate her doubt. Questions like "Might we not be completely mistaken about the external world?" must be regarded as nonsense, or

---

[25] nature of the Forms depend, either logically or really.

[26] Moore (1959), Ch.III, §50.

[27] The latter claim of appearance, not reality, is part of Leonard Woolf's justification of the habits of "*altifrons aestheticus*" in "Hunting the Highbrow." See Woolf (1970).

[28] One important difference is that on Carnap's view, *e.g.*, one can no more say that the world exists independently of our actual beliefs about it than one can raise the sceptic's doubt. But for each and every thing in the world the claim can be made intelligibly.

as implicitly pragmatic questions about which linguistic framework to employ. In neither case can they express the doubt that the Cartesian sceptic wants to express.[29] Ayer reaches a similar conclusion, drawing on the weak form of the principle of verifiability:

> We say that the question that must be asked about any putative statement of fact is ... Would any observations be relevant to the determination of its truth or falsehood? And it is only if a negative answer is given to this ... question that we conclude that the statement under consideration is nonsensical.[30]

But the worry of the sceptic cannot get off the ground here, since no series of sense-experiences could give evidence for believing that we were deceived about that same sense-experience as a whole. So, says Ayer, "anyone who condemns the sensible world as a world of mere appearance ... is saying something which, according to our criterion of significance, is literally nonsensical."[31]

As the emphasis on sense-experience may remind us, the positivists championed the idea that empirical science was to stand as the epistemic paradigm. It was "The Scientific Conception of the World ..."[32] that announced the programme of the Vienna Circle in 1929 with the goal of "*unified science.*"[33] Ayer summarized his own views thus:

> There is no field of experience which cannot, in principle, be brought under some form of scientific law, and no type of speculative knowledge about the world which it is, in principle, beyond the power of science to give.[34]

Science, for Ayer, is the paradigm of knowledge, and we must not be deluded by the belief "that there are some things in the world which are possible objects of speculative knowledge and yet lie beyond the scope of empirical science."[35]

---

[29] See Carnap (1956), p.207f.

[30] Ayer (1952), p.38.

[31] *Ibid.*, p.39.

[32] Hahn, Neurath and Carnap (1973), p.299.

[33] *Ibid.*, p.306.

[34] Ayer (1952), p.48.

[35] *Ibid.*, p.48. It might be tempting to suppose that a coherence-theory of justification plays some role in promoting scientism. (Thanks to Bruce Hunter for this suggestion.) Both Carnap and Neurath took coherence very seriously. For Neurath's views see Neurath (1983), pp.91-99. It is, of course, Neurath's metaphor of the ship at sea (see *ibid.*, p.92) that Quine regularly invokes when emphasizing his own coherentism. Ayer, as well, despite accepting a principle of verifiability, seems to have a coherentist bent:

Lest the example of logical positivism seem insufficient to show that scientism does not require realism--Seller treats positivism as a *form* of realism--one more instance should do the trick. What barrier stands in the way of being a *relativist* and also embracing scientism? After all, if truth and rationality are relative to cultures, then the thesis of scientism is true-for-a-scientistic-culture!

Should we conclude then that there is no connection between realism and scientism at all? I have tried to show that there is no logical connection. But there might be reasons to think that there is a contingent, historical connection between scientism and realism. Indeed, I think that two trends of thought apparent in the nineteenth and twentieth centuries display just this contingent connection: one is an important strain in recent analytic philosophy that has inherited its scientism from the logical positivists, but which is metaphysically realist; the other is an influential trend in Marxism that goes back to the nineteenth century and to which critical theorists such as Seller and MacKinnon find themselves responding. A thorough study of the frequency and strength of these connections would carry me far beyond the present project into a comprehensive history of ideas and perhaps into the realm of the sociology of knowledge. Thus, the remarks that follow are impressionistic, but I believe that they are suggestive of a deeper explanation for the ethical-political attraction of relativism.

III. Realism, Scientism and Analytic Philosophy: Contingent Connections

The principle of verifiability was the centre-piece of logical empiricism. An important doctrine that often accompanied it held that statements about morality, aesthetics and metaphysics possessed no truth-values--that they were literally nonsensical. The doubts of the Cartesian sceptic, as we saw, were considered to be of this order.

The pervasion of such so-called "nonsense," cried out for some positivistic explanation, which for ethics and aesthetics came in the forms of emotivism and prescriptivism. Metaphysical assertions were held to be confusions arising from the improper analysis of language or implicit recommendations for the adoption of new linguistic

---

[35] When one speaks of hypotheses being verified in experience, it is important to bear in mind that it is never just a single hypothesis which an observation confirms or discredits, but always a system of hypotheses. (Ayer, 1952, p.94.)

"[W]e define a rational belief," he says, "as one which is arrived at by the methods which we now consider reliable" (*ibid.*, p.100).

However, it seems to me that what is at work here is a *particular conception* of what coherence amounts to--a *scientistic* conception of coherence, according to which coherence is vitally connected with the usefulness of a system of hypotheses for the prediction of experience.

> [W]e test the validity of an empirical hypothesis by seeing whether it actually fulfils the function which it is designed to fulfil. And we have seen that the function of an empirical hypothesis is to enable us to anticipate experience. (*Ibid.*, p.99.)

frameworks. Other kinds of knowledge-claims, *e.g.* in the human or social sciences, were pressured to fit the prevalent philosophical conception of practices in the natural sciences, especially physics. The *unity of science* dictated that such disciplines be able to share methods of verification thought to be specifiable for the natural sciences. The "work of the Vienna Circle," included "creating the often neglected 'cross-connections' between the individual sciences so that it is possible to relate the terms of each science to every other science without effort,"[36] Neurath wrote. So, psychology was pushed toward an exclusive behaviourism--"behaviouristics"[37] in Neurath's jargon--while social scientists were to look for the "covering laws" required to explain historical events. Sociology, to borrow again from Neurath, was to be treated as "social behaviourism."[38]

All of these concerns, it should be clear, stem from treating all and only scientific knowledge as knowledge. Logical positivism has since met its official downfall, crushed between the two dogmas of empiricism, as Quine was to call them--reductionism and analyticity.[39] But its influence was considerable, and even its most respected critic remains a devotee of "epistemology naturalized" and "the science game,"[40] though he acknowledges "other good language games such as fiction and poetry"[41] and understands metaphor better than many.[42] As we saw in Chapter 3, Quine's privileging of physical science is partly to blame for his doctrine of the indeterminacy of translation.

In the wake of such an intellectual movement it should come as no surprise that, though the formal commitments of theory disappear, the temptation to elements of that older view linger on. Philosophers are creatures of habit as much as is anyone. And it is partly a reflection of these habits, I think, that has led to a revival of metaphysical realism and with it the correspondence-theory of truth. Beyond mere habit, we can also detect some impetus toward metaphysical realism and correspondence arising contingently from the rejection of the logical-empiricist dogma of reductionism. For this rejection signalled a shift away from "enumerative induction" toward abduction or "the inference to the best explanation" as an account of scientific reasoning.[43] If one is not prepared to reduce the statements of theory to, *e.g.*, statements in the language of sense-data or physicalism, then one faces questions about the ontological status of unobservable, theoretical entities. A non-instrumentalist account of the truth of scientific statements and theories requires justification for

---

[36] Neurath (1983), p.98.

[37] *Ibid.*

[38] *Ibid.*, p.71.

[39] Quine (1980), pp.20-46.

[40] Quine (1990), p.20.

[41] *Ibid.*

[42] Quine (1979).

[43] See, *e.g.*, Harman (1965).

believing in the existence of such entities, and, as Harman pointed out, "the inference from experimental data to the theory of subatomic particles certainly does not seem to be describable as an instance of enumerative induction."[44] Explanatory coherence thus acquired a new importance, to the point at which even correspondence-truth began to be justified on grounds that the (approximate) correspondence-truth of scientific theories gave the best explanation for the "success" of science.[45] The idea of correspondence to a reality whose nature and existence are independent of our epistemic capacities--I suggest--makes it possible at this stage of intellectual history for thinkers to retain the standard of science as arbiter of knowledge-claims.

Consider moral claims. Having abandoned positivism, can we now think of moral claims as having truth-values? If we want to, it seems that either we must admit that they are, not nonsense, but *all false*, or we must show how to give a naturalistic account of "moral properties," preserving the status of science as, if not the final judge of all epistemic claims, at least the epistemic protection-racketeer.

Take the former horn of the dilemma. If we are to make true moral statements, then there must be objective moral properties attaching to situations, events or persons. Moreover, we need some way of acquiring knowledge of such properties. But, according to J. L. Mackie's "argument from queerness," neither requirement is easily fulfilled:

> If there were objective values, then they would be entities or qualities or relations of a very strange sort, utterly different from anything else in the universe. Correspondingly, if we were aware of them, it would have to be by some special faculty of moral perception or intuition, utterly different from our ordinary ways of knowing everything else.[46]

Thinking of truth as correspondence to the metaphysical realist's world pressures us to have ostensible things or features of things to which we can point for confirmation of our moral claims as we would point to observable happenings, including instrument-readings, as confirmation of scientific hypotheses. But the natural sciences have found no such things or features, and given the motivational force of these things or features, it seems hard to imagine what they would be like--let alone how we could know of them, when our most comprehensive scientific theories show no trace of them. Thus, Mackie concludes, moral claims are all false, because moral terms lack any referent.

The second option--naturalism--is a response to the argument from queerness. What we need, says the moral naturalist, is an abductive argument for the existence of moral properties: we are justified in believing in objective values, provided those values offer the best--or at least a useful--explanation of some facet of human behaviour.[47] Such arguments, good or bad, take for granted the terms of adequacy

---

[44] *Ibid.*, p.90.

[45] See Putnam (1978), pp.18-33.

[46] Mackie (1977), p.38.

[47] See, *e.g.*, Sturgeon (1985) and (1986).

imposed by the scientistic credo. Naturalism in our era derives some of its credibility from scientism, and in the face of the metaphysical correspondence-theory of truth may seem the only viable alternative to subjectivism or the positivists' non-cognitivism.[48]

Lest it seem that I am ignoring the dominant trends of contemporary ethical theory, at least in North America, let me say something about contractarianism and constructivism. The existence of neither of these approaches to ethical theory threatens the link that I am proposing between scientism and metaphysical realism. Contractarian moral theory--at least in its most influential forms--takes *subjectivism* as its starting point. "[T]he theory of rational choice treats value as a subjective and relative measure, not as an objective and absolute standard."[49] By contrast, moral constructivism, popularized by Rawls,[50] recognizes that to say something interesting about moral theory one must forsake the idea of correspondence with a metaphysical reality. Indeed, Rawls has been championed recently by such liberal critics of metaphysical realism as Richard Rorty and Hilary Putnam.[51]

My point is that metaphysical realism and its correspondence-theory of truth, in a culture that values science and its authority, buoy up that authority by forcing candidates for non-scientific knowledge into a framework that mars their credibility. If we compare values to the entities or properties recognized by natural science, they do seem to be "queer" sorts of things. But this oddity arises from assuming that moral claims must be true or false in virtue of correspondence to a metaphysical world that has been impressively described by the natural sciences, where correspondence is lent credence by a scientistic notion of explanatory coherence. In the late twentieth century western scientism thrives on correspondence, but it has had other nutrients in the past. Here, then, is one reason for the relativist to mistake realism for the source of her worries.

A second reason is that the alliance of objectivism and scientism encourages an understanding of objectivity as "value-neutrality." Such neutrality has its predecessor in the positivists' expulsion of values from the realm of facts (or more importantly, from the realm of justification). This rigid fact-value distinction, as I suggested above, was preserved by the combination of scientism and realism. When we think of truth as metaphysical correspondence to the world, whose constituents are picked out by the natural sciences, we see truth as value-free, for the realist's world is independent of our epistemic capacities, and so, independent of our evaluative capacities. "For serious knowledge seekers," Lorraine Code observes, "objectivity and

---

[48] Attempts parallel to those of the ethical naturalist can be seen in the efforts of some linguists to reduce semantics to a component of universal grammar and of philosophers to reduce epistemic justification to "reliable processes," a position criticized in Chapter 2. Trying to treat reference as a (set of) physicalistic relation(s), another view I have already criticized (Chapter 1), manifests the same acceptance of scientistic terms of debate.

[49] Gauthier (1986), p.25.

[50] Rawls (1971).

[51] See Rorty (1991), pp.175-196 and PP3 302.

value-neutrality are virtually synonymous."[52] To the extent that our knowledge-claims seem value-laden, they will likewise seem further away from the truth, distorted by our own unjustifiable, subjective preferences. There is a clear separation between facts and values, and one possesses knowledge to the extent that one's views have been purged of the warping influence of such values. This idea of value-neutrality is not simply a consequence of realism; as we have seen, Platonism can be construed as a variety of metaphysical realism, and the Form of the Good is central to Plato's notion of knowledge. But when scientism is combined with metaphysical realism's strong notion of objectivity, values seem simply out of place.

A correlate of this treatment of objectivity as value-neutral is the treatment of knowledge as ahistorical, acultural and genderless. If truth is correspondence with a reality whose existence and nature are independent of our abilities to know about them, then we express the truth only to the extent that our ways of expressing it transcend our particular circumstances--our place in history, our broader culture, our position in society, our gender, our racial or ethnic background, and so on.[53] The relativist recognizes a problem in claims to value-neutrality and to a lack of, *e.g.*, cultural bias, but blames realism. This happens, I suspect, for two reasons: first, the relativist sees philosophical difficulties with metaphysical realism and attributes further problems to realism on the basis of an assumption about realism's "moral character," as it were. Secondly, perhaps scientism is so ingrained in our intellectual culture that the relativist does not recognize its distinct influence even as she criticizes "rational-scientific epistemology" or "the scientific imperative" (FMMS 527n23). Indeed, some recent work in environmental philosophy, far from confusing scientism and realism, actually embraces metaphysical realism as a way of ensuring that the natural world can be viewed as valuable in itself, quite apart from our evaluative or epistemic

---

[52] Code (1991), p.31. Code continues:

> The assumption of value-neutrality at the heart of the received view of objectivity owes much to the pride of place accorded to scientific knowledge, with *its* alleged value-neutrality, among human intellectual achievements. (*Ibid.*)

I find less sympathy, however, for her apparent claim that a strong conception of objectivity depends on, or is a response to, the threat of Cartesian scepticism (see *ibid.*, p.50). A similar claim is made in Bordo (1986). My argument in Chapter 2 was that Cartesian scepticism cannot get started without the metaphysical realist's conception of objectivity. However, it is a virtue of Bordo's work that scepticism is taken seriously in a way that it cannot be by the metaphysical realist. (My thanks to Edrie Sobstyl for pointing out this parallel.)

[53] The metaphysical realist's claim that there is but one true, complete theory or description of the world is also relevant here, but not as simply as it might seem. If one is a sceptic, then one will not claim to know the ideal theory. If one is a fallibilist, then one will be modest about one's epistemic claims--unless one is tricked into betraying fallibilism by, *e.g.*, the success of science and the assumption of a fact-value dichotomy, the objects of my criticisms here.

capacities.[54] However, these proposals concerning the relativist's reasoning, especially the latter proposal, would require considerably more study than I can devote to them here.

This whirl-wind summary is unlikely to convince the unsympathetic. But I hope that it does show how the politically motivated relativist might come to believe that realism is a position that she should oppose. If she assumes the dichotomy of realism and relativism, then her choice of theoretical rhetoric will be clear. It is also worth noting that my own position differs from metaphysical realism at precisely the points at which, I have suggested, metaphysical realism is vulnerable to the influence of scientism. (This, of course, is no mere coincidence.) My position is anti-scientistic in a number of ways. As my arguments of Chapters 3 and 4 suggest, it is compatible with recognizing many alternative descriptions of the world, including practically incommensurable ones--though it does not rule out questions about which interests and practices are worthwhile. As well, by abandoning the metaphysical realist's conception of objectivity and the correspondence-theory of truth, my view is not subject to the pressure to think of values as natural properties of the (meta-) physical world in the way that the argument from "queerness" suggests. Indeed, viewing truth as internally related to warranted assertibility blurs any rigid fact-value distinction.

Also--but more tentatively--the recognition that meaning is normative, that semantics cannot be naturalized, leads us in the direction of seeing linguistic meaning as connected primarily with linguistic use, and hence, away from thinking of meaning as given primarily by the "literal" significance of declarative sentences. With this levelling of semantic priorities goes a levelling of the privileged status of fact-stating discourse, so amenable to expressing the *results* of scientific investigation and encouraging a picture of knowledge as correct representation, rather than as akin to a variety of abilities. Moreover, the rejection of naturalized semantics is itself an anti-reductionist conclusion, acknowledging that the search for empirical regularity and predictive power is just one paradigm and cannot be usefully applied in all areas of inquiry and understanding.

Finally, understanding meaning as use enables the recognition of biases of culture, era, gender and so on, since what can intelligibly be said or thought faces the practical limitations of historical and socio-cultural context. These points I raised in Chapter 2, but it is appropriate to repeat them, if only as enticement for the relativist.

IV. Realism, Scientism and Marxism: More Contingent Connections

Another likely source of the relativist's conflation of realism and scientism is traditional Marxist theory. Marxism meets relativism at two points, both related to the role of ideology in Marxist theory.

(1) Marx's later writings are clearly realist, as are the writings of figures like Engels and Lenin. As well, the dominant strains of Marxist theory, and Marx's own views, are scientistic. But it is most interesting that historical materialism, especially in versions that emphasize ideology and false consciousness, merges realism and

---

[54] This is much the point of Stan Godlovitch's "acentric" account of value in Godlovitch (1992). For comparison, see also Noel (1989).

scientism in functional explanations of the development of class relations in history--and functional explanations are anti-subjective. A predictable reaction is to emphasize the subjective end of the objective-subjective dichotomy. Let me explain this point.

Marx's realism is evident in his historical *materialism*, described in *Capital* as an inversion of Hegel's dialectical idealism (MEW XXIII 27).[55] Consider this passage from *The German Ideology* in which he and Engels give an early account of the materialist conception of history:

> The premises with which we begin are not arbitrary, not dogmas; they are real premises from which one can abstract only in imagination. They are real individuals, their activity and their material conditions of life, both as these conditions are found and as they are produced through the activity of individuals. These premises are thus verifiable [*konstatierbar*] in a purely empirical way. (MEW III 20)[56]

Not only are the "real" and the "material" emphasized here in contrast to the "ideal," but the relevance of science is suggested by the reference to empirical verification. Indeed, Marx is fond of viewing his methods as scientific. "There, where speculation ends--with real life--real positive science begins ..." (MEW III 27).[57] Elsewhere he compares his investigations to those of the physicist (MEW XXIII 12)[58] or the evolutionary biologist (MEW XXX 578)--Marx's correspondence is full of references to Darwin--and in a passage from the Preface to *A Critique of Political Economy* he writes of the conflicts that arise between the productive forces and the relations of production in a way that ostensibly sets his method apart from the falsehood of ideology:

> [O]ne must always distinguish between material upheaval in the economic conditions of production, which can be verified in accordance with natural science, and the legal, political, religious, artistic or philosophical, in short, ideological forms, in which people become conscious of these conflicts and fight them out. (MEW XIII 9)[59]

Likewise, science and technology are to revolutionize the forces of production so that "society's necessary labour" is reduced to a "minimum,"[60] leaving individual talents to flourish in creative activity, rather than stagnate in the satisfaction of basic needs.

The importance of both science and realism have continued to be emphasized by influential theorists in the Marxist tradition. In his

---

[55] See Marx (1967), I, p.19.

[56] See Marx and Engels (1970), p.42.

[57] See *ibid.*, p.48.

[58] See Marx (1967), I, p.8.

[59] See Marx (1977), p.389f.

[60] Marx (1953), p.593. Quoted in Habermas (1971), pp.48-50. Habermas emphasizes this passage as evidence of Marx's scientism.

Materialism and Empirio-criticism Lenin castigates the "makhistuy," Russian followers of Ernst Mach's positivism, for their "idealism" (which he thinks is indistinguishable from Berkeley's).[61] And Louis Althusser (in a passage quoted by MacKinnon), despite a departure from Marx's analysis of ideology, retains the distinction between ideology and science: "We know that a 'pure' science only exists on condition that it continually frees itself from ideology which occupies it, haunts it, or lies in wait for it."[62] To the extent that radical theorists write partly in response to the long tradition of Marxist theory, it should not be surprising that a rejection of Marxist scientism should issue forth in a rejection of realism. Provided the realist-relativist dichotomy is at work, then, relativism may seem once again to be the appropriate radical stance.[63]

What is especially interesting about Marx's (and Marxism's) combination of realism and scientism, however, is its functional explanations of historical development. Marx's conception of ideology, when wedded to his overall theory of history, offers us explanations of why particular ideas are the "ruling ideas" (MEW III 47)[64] of an era--explanations of the form "the cause occurred because of its propensity to have that effect ..."[65] That is to say, the "ruling ideas" are believed by a significant number of people because it serves the interests of the ruling class that those ideas be found credible. Such functional accounts of the social genesis of belief, if taken to an extreme, tend to rule out entirely the validity of individuals' own self-interpretations. If one asserts that only the theorist's interpretation of the experience of the oppressed has a purchase on the truth and that all else is false consciousness or ideology, then one has taken the objectivist side of the objective-subjective dualism right into the sphere of self-knowledge.[66] This tendency obtains influential expression in two strains of theory

---

[61] Lenin (1927), p.5f.

[62] Althusser (1979), p.170. See FMMS 527n23. I discuss Althusser's views at greater length below and in Chapter 8, §III.

[63] MacKinnon says that she is influenced by the Marxism of Georg Lukacs whose position is "more hospitable to feminism" than that of the "dominant tradition" because the former "embrace[s] its own historicity" and thereby "see[s] the scientific imperative itself as historically contingent" (FMMS 527fn23). But as I have said, I hesitate to describe MacKinnon as a relativist; nor does she describe herself as one. Her relevance here lies in her conflation of realism and scientism, a conflation whose occurrence amongst critical theorists, I am suggesting, leads to relativism in some cases.

[64] See Marx and Engels (1970), p.64.

[65] Cohen (1978), p.281.

[66] I shall explore this theme in Chapters 7 and 8.

influenced by Marx and Marxism, the critical theory of the Frankfurt School and the structuralist Marxism of Louis Althusser.[67]

The former case can be represented by Herbert Marcuse, who in *One-Dimensional Man* warned that "advanced industrial civilization" had developed "new forms of control" that ensured that "[a] comfortable, smooth, reasonable, democratic unfreedom prevails ..."[68] Central to his assessment of advanced industrial culture was the thesis that technological society produces a condition in which "individuals identify themselves with the existence which is imposed upon them and have in it their own development and satisfaction."[69] By offering to enough people the conveniences of technological innovation, advanced industrial society effects a kind of ideological control that shapes the very desires of its subjects, discouraging the growth of criticism and stifling opposition by satisfying and then reproducing those desires. Social reality itself becomes ideological, signalling the end, not of ideology, as proclaimed by some of Marcuse's contemporary opponents, but rather, of reality. One-dimensional "men," in Marcuse's exclusionary phrase, are "swallowed up" by their "alienated existence" and identify with those needs and desires that serve to maintain the *status quo* of their own subordination to technology and capital.[70]

Even if one recognizes some truth in Marcuse's analysis, the pervasiveness that he attributes to these new forms of ideological control is likely to provoke two related responses: indignation at the perceived portrayal of the oppressed as helpless dupes, and concern that Marcuse has defeated his own purposes by "portray[ing] oppression in its full force, as inescapable ..." (SAAO 501). This in turn is likely to drive the dichotomizing thinker into the arms of relativism.

In Althusser's case a similar pair of problems is often at hand. They grow from slightly different roots, however. For Marcuse, the threat to subjectivity arises from a new kind of ideology, which, on Habermas's account of Marcuse, differs from the old kind in two ways.[71] First, it presupposes that class-antagonism, which was controlled and masked by the old ideology of 19th-century capitalism has been *recognized*, so that the old ideology has been rendered

---

[67] As Bhiku Parekh observes, Althusser claims that "a worker is better able to understand Marx than [is] a 'bourgeois professor'" (Parekh, 1982, p.173). Althusser writes,

> It is paradoxical that highly 'cultivated' intellectual specialists have not understood a book [*Capital*] which contains the Theory which they need in their 'disciplines' and that inversely, the militants of the Workers' Movement have understood this same Book, despite its great difficulties. (LP 73; *Cf.* LP 16, 68ff)

[68] Marcuse (1964), p.1.

[69] *Ibid.*, p.11.

[70] Similar themes can be found in feminist theory. See, *e.g.*, MacKinnon (1987), p.54. See also Bartky (1982).

[71] Habermas (1970a), pp.81-122.

ineffective. Recent capitalism with its "technocratic consciousness"[72] cannot afford to indulge in the same degree of unbridled oppression and exploitation (at least in the "North Atlantic" and Oceanic democracies), and so, must make concessions to the oppressed without relinquishing power.[73] Nineteenth-century socialism played the Ghost of Christmas-Yet-to-Come to capitalism's Scrooge, and capitalism has reacted accordingly. Its new ideology must mask the fact that, while Scrooge keeps Christmas in his heart, an unjust distribution of power remains. Secondly, the "new" ideology operates "with the aid of rewards for *privatized needs*."[74] The oppressed learn to enjoy their oppression and find fulfilment in social and political inequality.

For Althusser, however, the all-pervasiveness of ideology stems from theoretical difficulties, rather than cultural pessimism. His work shows a peculiar tension between his acceptance of a sharp distinction between science and ideology and his view that "*ideology has the function (which defines it) of 'constituting' concrete individuals as subjects*" (LP 171). Like Marx and the Frankfurt school, Althusser draws a distinction between knowledge (Marxist science) and ideology. But whereas Marx might be taken to hold that ideology is an abuse of abstraction--a depiction of contingent and local features (real or imagined) of the human world as necessary and universal, in a way that benefits the ruling class[75]--Althusser adds his own twist. "Ideology," he says, is a "necessarily imaginary distortion" that represents "the (imaginary) relationship of individuals to the relations of production and the relations that derive from them" (LP 165). The "imaginary" marks Althusser's adoption of the conceptual apparatus of the psychoanalytic theory of his colleague, Jacques Lacan.

In his *Écrits* Lacan offers an interpretation of Freudian psycho-sexual development that invokes the notion of a "mirror-stage" in that development. "The child," he claims, "at an age when he is for a time ... outdone by the chimpanzee in instrumental intelligence, can nevertheless already recognize as such his own image in a mirror."[76] This alleged recognition of self, Lacan seems to suggest, is analogous to the "I's" immediate consciousness of itself as pure thought in the Cartesian *cogito*.[77] However, the child identifies, not with *res*

---

[72] *Ibid.*, p.112.

[73] Chomsky argues a similar point, but without Marcuse's excesses. He also observes that contemporary capitalism still depends crucially on violent repression, but the violence is exported to non-industrialized nations, whose resources fuel the relatively affluent economies of the "West." These themes pervade Chomsky's political writings. See, *e.g.*, Chomsky (1989) and Chomsky (1993).

[74] Habermas (1970a), p.112.

[75] There is ample evidence for this reading in *The German Ideology*. Althusser's reading of that work (LP 158-160) fails to engage the text on any more than a superficial level.

[76] Lacan (1989), p.502. In what follows I have been aided by Eagleton (1991), pp.136-146, and by Rose (1982).

[77] I shall discuss Descartes' views briefly in Chapter 6.

cogitans, but with a mirror-*image*, an "*imaginary*" representation of itself. Such simple identification masks "the turbulent movements that the subject feels are animating him,"[78] the desires and sensations not displayed in the mirror. This "primordial form" in which the "*I*" is "precipitated,"[79] then, is a distortion, but it constitutes for the child "the model and basis for all its future identifications."[80] It is this fiction of simple self-unity that "situates the agency of the ego, before its social determination."[81]

In a similar sense Althusser takes ideology to involve an "imaginary distortion." In ideology individuals "misrecognize" themselves, finding their identity automatically and unreflectively in the existing social order. The existence of subjects and the existence of the current regime are as *one*, and the subject can envisage the persistence or cessation of one only with the persistence or cessation of the other. Thus, she experiences her own worth and sense of "freedom" in her subordination. Moreover, without ideology there would be no subjects, since "ideology has the function (which defines it) of 'constituting' concrete individuals as subjects." Individuals are constituted as subjects by an ideology that "interpellates" or "hails" them, as though each were a free, conscious agent, responsible for her actions, able to do as she wants, a "centre of initiatives" (LP 182). But, as with Lacan's "mirror-stage," this simultaneous "situation" of the "agency of the ego" and "imaginary" awareness of self, according to Althusser, masks the unconscious factors influencing the subject's "constitution"--*viz.*, the subordination of the subject to capital and the mechanisms, ideological and other, needed to reproduce the existing relations of production.

Now, Althusser distinguishes between particular ideologies and ideology in general. And while particular ideologies may come and go, "ideology *in general has no history*" (LP 160f). It exists "outside" history in the sense that historical events are always structured and influenced, in part, by ideology. Ideology, says Althusser, is "an *omnihistorical* reality" (LP 161), and "there is no practice except by and in an ideology" (LP 170). Paradoxically, then, the very thing that enables purposeful, reflective action (practice)--that which "constitutes" individual human organisms as "subjects"--also makes the subject systematically self-ignorant, since it oversimplifies and distorts the relationship of individuals to the relations of production. And this holds for a proletarian ideology--which Althusser must countenance, given the "omnihistorical reality" of ideology--as much as for bourgeois ideology.

But a deeper worry about self-ignorance lurks here, as well, because it would seem that Althusser's dichotomy of science and ideology prevents him from making sense of the idea of a reflective, critical practice, opposed to the ruling ideology. If practice requires an ideology, then a practice of resistance requires an ideology of resistance. However, it is implausible to suppose that there could be

---

[78] Lacan (1989), p.504.

[79] *Ibid.*, p.503.

[80] Rose (1982), p.30.

[81] Lacan (1989), p.503.

a reflective practice of resistance without some kind of *critique* of the ruling ideology--and that is the role played by Marxist *science* in Althusser's scheme. But if science remains opposed to ideology, and if ideology is necessary for practice, then there can be no reflective activity of resistance. Althusser's account, as Ted Benton says, "leaves no theoretical room for a discourse and practice of ideology which *resists and opposes*"[82] the machinations of the ruling ideology. Althusser has, Benton suggests, confused an analysis of the ruling ideology with an analysis of ideology in general, and the consequence is that each subject is seen to experience her own worth and sense of "freedom" in her own subordination.

It appears, then, that Althusser's position also portrays oppression as inescapable (except, maybe, by luck) and the oppressed as dupes.[83] If ideology constitutes individuals as subjects, if "*individuals are always already subjects*" (LP 172), if "there is no practice except by and in an ideology," and if ideology is an "imaginary distortion," then the chances of freedom and self-knowledge are greatly limited, even ruled out. It is a mystery how one even reaches the "subject-less discourse" (LP 171) of science to which MacKinnon objects.

It may seem a paradoxical suggestion that radical thinkers might be tempted by relativism in response to Althusser, because Althusser is sometimes accused of *being* a relativist himself.[84] But it is worth remembering that relativism is just another kind of metaphysical realism, and Althusser's position--relativist or not--emphasizes a kind of sceptical metaphysical realism about the self, according to which I could be completely mistaken about myself, according to which my capacity to act *depends* upon my self-ignorance, because as a subject I am constituted by ideology. If Althusser's theorizing is part of Marxist "science," then its metaphysical realism about the self presents science and realism in close alliance. And if metaphysical realism can be confused with internal realism, and if "realism," like Marx's and Althusser's *materialisms*, is thought of as opposed to *idealism*, then the alliance will seem that much stronger.

It is in response to her perception of a similar tendency in feminist theory that Seller rejects rational-scientific epistemology:

> [T]he political problems of the rational-scientific epistemology are made more acute when questions of ideology and false consciousness are introduced. At best, the use of this epistemology appears to be profoundly undemocratic. At worst, it is an exercise in domination. At best, some women are telling other women what they are like, what their interests are, and how they might best be served. At worst, some women are imposing their own interests on the movement as a whole. (RVR 172)

Once again, to the extent that Seller conflates realism and scientism and sees no middle ground between realism and relativism, she is led to conclude that ideology and false consciousness are useless notions,

---

[82] Benton (1984), p.105.

[83] See *ibid.*, p.107 for this same criticism.

[84] See, *e.g.*, RTH 158-160.

realism must be abandoned, and relativism offers the only chance of a "democratic epistemology" (RVR 169). But the conflation of realism and scientism is made easier by the long-standing concurrence in Marxist-influenced radical theory of both realism and scientism.

What is less clear is whether Jean Grimshaw, who bears the brunt of Seller's criticisms, is guilty of the scientism that Seller sees in her use of the ideas of ideology and false consciousness. The examples that Grimshaw cites of applications of *Ideologiekritik* in feminist thinking are just the scientific critiques of cases of sexist thinking that Seller troubles herself about whether or not to accept, trapped by her assumption that science must be all or nothing to epistemology. Grimshaw is careful to point out that "There is no consensus about what is in fact in women's interests," though she adds that it is a starting assumption of feminism that some things are "*not* in women's interests."[85] To say this requires something more substantial than relativism, but not *metaphysical* realism:

> The questions 'Whose interests are served by these beliefs?' or 'What *difference* does the holding of these beliefs make to the social situation of women?' ... require that it be possible, in some sense, to talk about what is 'objectively true' about the situation of women. But this is not a notion of objectivity which implies a sharp split between 'facts' and 'values', or the possibility of a totally detached theoretical stance.[86]

Grimshaw's endorsement of this conception of objectivity raises the possibility that one could be an internal or pragmatic realist and still make use of a critical conception of ideology. Indeed, I see no reason why one could not hold that particular beliefs, practices or institutions in particular cases merited functional explanations in terms of the interests served by the presence of such beliefs, practices or institutions without thereby being committed to an all-encompassing theory of history and the social determination of belief. This kind of local, piecemeal analysis might require that people could be mistaken in their self-interpretations, but that is surely true.[87] And it would require neither depicting them as mere dupes, nor depicting oppression as all-pervasive and inescapable. This option seems unavailable, only if one thinks that the place of "rational-scientific epistemology" is either everywhere or nowhere, and this dichotomy arises from conflating scientism with metaphysical realism and from the prior dichotomy of realism and relativism.[88]

---

[85] Grimshaw (1986), p.101.

[86] *Ibid.*

[87] Or so I shall argue in Chapters 7 and 8. Seller seems to agree on this point--that is part of what is supposed to distinguish relativism from subjectivism on her account. See RVR 180, 182.

[88] Seller's criticisms of Grimshaw have echoes of an exchange between Winch and MacIntyre over the nature of social science. She cites Winch--and Wittgenstein--at RVR 179n6, and criticizes a paper by MacIntyre at RVR 177f. However, Winch's position is more sophisticated

(2) So much for the first dimension of Marxist influence on relativism. The second is related to the first, taking its cue also from the distinction between science and ideology. This time relativism comes as a response to a criticism of traditional Marxism.

The science-ideology distinction separates not only truth from falsehood and reality from illusion. It also separates two kinds of portraits of the social world: those which distort that world to make the interests of a minority appear continuous or identical with the interests of society as a whole, and those which only minimally distort that world, because they represent the interests of the oppressed majority, which on the sheer basis of numbers, are closer to continuity with the interests of society as a whole. Marx and Engels sketch such a picture in *The German Ideology*:

> For each new class which places itself in the situation of one ruling before it, it is necessary, in order even to carry out its purpose, to represent its interest as the common interest of all members of society, *i.e.*, ideally expressed; in order to give the form of universality to its ideas, it represents them as the only reasonable and generally valid ones. (MEW III 47)[89]

Because the proletarian movement, according to Marx and Engels, "is the independent movement of the immense majority in the interests of the immense majority" (MEW IV 472f),[90] the acquisition and maintenance of its power depends less on ideological distortions: once power is equitably distributed a minority of recalcitrant capitalists, say, will be less of a threat to that power. To the extent that distortion is needed, it will be distortion of a lesser degree than those of bourgeois ideology, for the reason that proletarian thought will already represent the interests of the "immense majority." Marx and Engels even speculate that with the end of class-divisions under communism, ideology will fade away altogether:

> This whole appearance, that the rule of a certain class is only the rule of certain ideas, comes to an end, of course, as soon as the form of social organization in general ceases to be the rule of a class, as soon, therefore, as it is no longer necessary to represent a special interest as a general one or "the universal" as ruling. (MEW III 48)[91]

---

[88] than Seller's, and MacIntyre's clings to the paradigmatic status of science, albeit in a subtle way. See Winch (1958) and (1970), as well as MacIntyre (1970a) and (1970b). MacIntyre is correct to insist on the possibility of applying terms of ideological critique (MacIntyre, 1970b, p.118), but Winch is also correct to criticize the scientism implicit in MacIntyre's position. See Winch (1970), pp.94-111.

[89] See Marx and Engels (1970), p.65f.

[90] Marx (1977), p.230.

[91] See Marx and Engels (1970), p.66.

This rosy vision of the future, however, is called into question by doubts about the inevitability of communism or the claim of critical theorists to represent the interests of the oppressed. A distortion that is claimed to be minimal is still a distortion; rather than pretend that one group has special access to the truth, why not settle for studying the sociology of knowledge? So runs this criticism.

Karl Mannheim describes the process in historical terms as the evolution from a "special formulation" of the "total" conception of ideology to a "general" formulation:

> As long as one does not call his own position into question but regards it as absolute, while interpreting his opponent's ideas as a mere function of the social positions they occupy, the decisive step forward has not yet been taken. ... In contrast to this special formulation, the general form of the total conception of ideology is being used by the analyst when he has the courage to subject not just the adversary's point of view but all points of view, including his own, to the ideological analysis.[92]

Mannheim's "total" conception of ideology is to be contrasted with the "particular" conception. The total conception portrays the thought under study as part of a complete "thought-system" whose content and "conceptual framework" is to be subjected to a "functional analysis" so that it may be understood as "an outgrowth of the collective life" of its holders.[93] The particular conception, by contrast, concerns itself only with the content of specific assertions and offers an "analysis of ideas on a purely psychological level," showing how "this or that interest" of an individual "is the cause of a given lie or deception."[94] According to Mannheim, it is the total conception of ideology that is relevant to Marxism, but Marxism does not take the "decisive step forward" to the general formulation of the total conception of ideology, by whose emergence "the simple theory of ideology develops into the sociology of knowledge."[95]

It should be clear, however, that such a move brings us very near to some version of the Strong Thesis discussed in Chapter 4, and that, of course, entails epistemic relativism. Mannheim thinks he can avoid this consequence, but it is possible to see how the temptation of relativism might be felt at this point, if one is sympathetic to the idea that thought is socially determined.[96] Such temptation can take

---

[92] Mannheim ([1936]), p.77.

[93] *Ibid.*, p.57f.

[94] *Ibid.*

[95] *Ibid.*, p.78.

[96] This is a thesis to which Hesse urges ST-proponents to soften their commitment. The sociology of knowledge should be concerned to show how beliefs can be causally influenced by social contexts, but not completely determined, she thinks, because she takes such determinism to be incompatible with freely choosing, or being rationally persuaded to accept, the Strong Thesis. See STSS 50.

at least two forms. First, one may sympathize with the advocates of the Strong Thesis and regard the sociology of knowledge as a field of "disinterested research," as Barnes and Bloor seem to. This option appears consonant with Mannheim's views. But one might also want to respond to Mannheim's criticism by finding a way of taking the "decisive step forward" without abandoning critical theory. Indeed, there are passages from Marx and Engels that might be taken (though I think incorrectly) to support such a "decisive step"--*e.g.*,

> [W]e start from real active people, and from their real life-process we show the development of the ideological reflexes [*der ideologischen Reflexe*] and echoes of this life-process. The hazy formations [*Nebelbildungen*] in the human brain are also, necessarily, sublimates of their material life-process, which is empirically verifiable and tied to material premises [*Voraussetzungen*]. Morality, religion, metaphysics, as well as other ideology, and their corresponding forms of consciousness herewith no longer retain the appearance of independence. They have no history; they have no development, but people, developing their material production and their material intercourse, thereby change their reality as well as their thinking and the products of their thinking. Consciousness does not determine life, but life determines consciousness. (MEW III 26f)[97]

Two related points might be extracted from this passage. First, the claim that consciousness is determined by life, and seemingly consists of "ideological reflexes and echoes" gives the sense that cognitive activity is but an epiphenomenon of material conditions, the hum and exhaust of history's engine. This impression is reinforced by Pascal's translation of '*Nebelbildungen*' as 'phantoms' (rendered here as 'hazy formations'). On this view, the truth and falsity of beliefs would be irrelevant to their acquisition, due to the complete determination of intentional attitudes by non-rational, historical circumstances. Any useful study of beliefs would be grounded, not in anything like traditional western epistemology, but in a sociology of knowledge.

On the other hand, we could be enticed by another way in which forms of consciousness might have no development. A plausible conclusion might be that the conceptual systems generated by any one historical set of material conditions are *closed* systems, whose concepts cannot be understood from one historical epoch to another, but within which it is useful to distinguish "true" beliefs from "false" beliefs. The thought of historical eras is contained by incommensurable conceptual schemes, such that the thinkers of one era concern themselves with entirely different topics from those of other eras, despite apparent similarities. So, rather than toward epistemic relativism, this interpretation gravitates toward conceptual relativism. This view I have already discussed and found wanting.

To the extent that passages like the one quoted above might tempt some interpreters to find either epistemic or conceptual relativism in Marx's thinking, it might be thought that one or both of these views has a role to play in a viable critical theory. This, I submit, is a mistake, but we see here the temptation of relativism for radical

---

[97] See Marx and Engels (1970), p.47.

theory. Concerned that Marxist objectivism does not subject itself to the same critical standards demanded of the positions that it criticizes, the relativist turns in the direction of subjectivism.

I have been offering some explanatory hypotheses concerning the influences that may lead a critical theorist toward relativism, and I have tried to provide the relativist with hints as to why relativism is not needed to satisfy the concerns of radical theory. This discussion of relativism and my arguments of earlier sections and of Chapters 3 and 4 have returned often to the problem of other minds, dressed, as I have said, in cultural garb. Having moved from the world to others, I now want to tighten the circle of discussion even further and focus on problems of self-knowledge. Here, too, the metaphysical realist's notion of objectivity exercises its influence. That this should be so has been hinted at in a number of ways. Functional and structural explanations of the determination of belief, as we have seen, seem to cast doubt on the possibility of self-knowledge. As well, Putnam's argument against metaphysical realism was based on the principle that a person cannot be completely mistaken about the meanings of her own words. And the relativist ran into difficulty when it became apparent, on the one hand, that conceptual relativism might pose a sceptical threat to self-knowledge, and on the other hand that epistemic relativism is ultimately indistinguishable from self-defeating subjectivism. So, let us bring knowledge of the self under scrutiny and try to discover how the neuroses that trouble the realist and the relativist display themselves yet again.

Chapter 6: Self-Knowledge: Descartes, Hume and Kant

I have been exploring the dichotomy of objectivity and subjectivity with a focus on knowledge of the "external" world and of other minds and cultures. The latter concerns have arisen naturally from the dialectic of realism and relativism. Relativism reacts to realism by stressing the subjective pole of the dichotomy, taking us away from "the world" to the many worlds of different cultures. But this descent from the peak of objectivism becomes a rapid fall into subjectivism.

The problems that our familiar dichotomy creates for knowledge of the "world" and of others take on a sharper form when we turn to a third classical issue of epistemology, the problem of self-knowledge. Here, it seems, the subjective pole of the dichotomy should be strong. What could be more subjective than the knowing subject? Do I not know the contents of my own mind better than anything else? Yet, the need for subjectivity without subjectivism drags us into the field of the objective pole, too. The self, we want to say, is real, not illusory, and we must stake out a place for its objective existence. But objectivity in the metaphysical sense raises sceptical worries, and the self threatens to slide behind a veil of deception and ignorance.

In this chapter, I consider some historical attempts to combine the objective and the subjective outlooks in understanding self-knowledge. Descartes and Hume, I submit, are caught in the oscillation between objectivity and subjectivity in their accounts of self-knowledge and its object. Kant weakens the dichotomy, but his views are tied to it, presupposing a sharp contrast between conceptual scheme and empirical content, and neglecting the relevance of language for self-knowledge.

I. Descartes: On Mixing the Subjective with the Objective

The curious form that the dualism of objectivity and subjectivity can take when confronted with problems of self-knowledge is manifest in Descartes' simultaneous endorsement of a kind of subjectivism and a kind of metaphysical realism about the self. This combination will seem implausible, but I think that it accurately describes Descartes' acceptance, on the one hand, of the infallibility of self-knowledge, and on the other, of metaphysical dualism, which entails that the soul is a thinking *substance*, which depends for its existence on no other thing, save the workings of an all-powerful, all-benevolent, all-knowing deity. Let us see how this might work.

The particular sort of epistemic neurosis that I want to diagnose in Descartes' work depends on attributing to him the combination of two theses: (i) that the self is a thinking *substance* whose existence is dependent upon nothing but (a) the creative and preserving influence of an all-powerful, all-benevolent, all-knowing deity and (b) that substance's capacity to know itself to exist as "pure intelligence";[1] (ii) such pure intelligence can be mistaken about neither the current

---

[1] Gueroult (1984), p.31.

occurrence of its own thoughts (ideas), nor the *contents* of those thoughts, provided it is sufficiently attentive.[2]

These two theses, or variations on them, fit with divergent readings of the details of Descartes' philosophy. So, in a sense, my diagnosis does not stand or fall with the particular points of interpretation I shall suggest. More important for my purposes is that these theses have structured Anglophone philosophers' conceptions of Descartes, especially in our century, and have contributed to the intellectual climate that makes us vulnerable to the philosophical ailments that I have been criticizing. Also, since my brief historical junket centres on extracting from the work of Kant some useful strands to be woven into my own account of self-knowledge in Chapters 7 and 8, the minimal constraint on my readings of Descartes and Hume is that they satisfy a plausible interpretation of Kant's *critique* of Descartes and Hume. I do, however, want to show that it is at least *prima facie* plausible to ascribe the mixtures of subjectivity and objectivity of which I speak.

Although both (i) and (ii) have influenced traditional conceptions of Descartes' thinking,[3] (i) is more easily gleaned from Descartes' writing. In the *Meditations* it is held to be an immediate consequence of the *cogito* that there is a *substance*[4] that is the ground of the pure intelligence which--in the *cogito*--knows itself to exist. Not only do I exist, but I am a thinking thing. For the self to be a thinking *substance* is for it to be "a thing capable of existing independently" (CSM II 30) of all *created* substances, but it "must be created by God in order to exist" and it cannot cease to exist without "God's denying his concurrence" to it (CSM II 10). This is an important step for the argument that emerges in Meditation VI for the distinctness of mind and body and for the immortality of the soul.

That the self is a thinking substance is a point made by Descartes in other places. In the *Discourse* he writes, "I knew I was a substance whose whole essence or nature is simply to think" (CSM I 127). In his *Principles of Philosophy*, he divides the universe into two kinds of created substance, "corporeal substance and mind (or created thinking substance)" (CSM I 210), and says that created substances "need only the concurrence of God in order to exist" (CSM 210).

So, a substance depends on no other created substance for its existence. *A fortiori* it would seem that the self can exist apart from the knowledge and, indeed, the epistemic capacities of other selves (though Descartes might have insisted that divine benevolence ensured the irrelevance of such contingencies). It is interesting to compare this objectivity and independence of the self with the objectivity and independence of the "external" world, characteristic of metaphysical

---

[2] While "metaphysical" doubt (the evil-demon hypothesis) is in effect in the *Meditations* the veridicality of these ideas is dubious. For the contrast between metaphysical doubt and "natural" or "psychological" doubt see Gueroult (1984), p.17f. and Schouls (1980), p.96.

[3] Curley suggests that an even stronger incorrigibility thesis than (ii) has come to comprise part of the English-speaking world's picture of Descartes, due largely to the influence of Ryle's attack on the "official doctrine" of the "Ghost in the Machine" presented in Ryle (1949), pp.11, 15f. See Curley (1978), pp.170ff.

[4] The term 'substance' does not appear until Meditation III.

realism. The only difference seems to be that for each self there is one epistemic agent (other than God), who does not stand in the same relation of independence from that self--namely, the self itself. In other words, without having examined the *degree* of Descartes' commitment to the infallibility of self-knowledge, we seem already to be able to conclude that for Descartes each self, *simply in virtue of being a Cartesian substance*, possesses a kind of privileged access to itself. Although I can doubt the existence of a (non-divine) pure intelligence distinct from myself, I cannot coherently doubt my own existence as a thinking thing, according to Descartes. So, with great dispatch we are confronted with issues of relevance to thesis (ii).

But how much closer to (ii) can we get than the presumed indubitability of the *cogito*? That Descartes is committed to (ii) is not as clear as the views of mainstream Anglophone philosophy would have it. As Curley observes, there are passages from Descartes' writings that might even seem to suggest the contrary,[5] though as he wittily remarks, "The topic of consciousness is certainly not one on which Descartes can be accused of a foolish consistency."[6] But without wishing necessarily to deny the contrary tendency to be found in Descartes, I think that some passages are quite suggestive of an acceptance of the infallibility of self-knowledge. Indeed, if there is a tension in Descartes' writing, then such tension serves as a further confirmation of my diagnosis, for it is the nature of an epistemic neurosis that incompatible theses must be embraced.

One passage that is quite suggestive of a commitment to (ii) appears in the Second Meditation:

---

[5] Curley cites three passages from (1) the *Discourse*, (2) a letter to Christian Huygens, and (3) *The Passions of the Soul*. The first is compatible with my version of the incorrigibility thesis:

> [M]any people do not know what they believe, since believing something and knowing that one believes it are different acts of thinking, and the one often occurs without the other. (CSM I 122)

Unless one takes time to reflect on one's beliefs, and on the ideas regarding which beliefs are judgments (see CSM I 205ff. and CSM II 41 for the connection between judgment and belief), then one may believe without knowing what she believes--I take Descartes to be saying here. But this is compatible with holding that one cannot be mistaken about the occurrence or content of one's thoughts (including judgments), if one has been sufficiently attentive in reflecting on them.

Similar remarks apply to the letter to Huygens and to the passage from *The Passions of the Soul*: "... experience shows that those who are most strongly agitated by their passions are not those who know them best ..." (CSM I 339). As Curley notes, a couple of paragraphs earlier Descartes seems to accept an incorrigibility thesis about the passions, which concludes, "...even if we are asleep and dreaming, we cannot feel sad, or moved by any other passion, unless the soul truly has this passion within it" (CSM I 338). See Curley (1978), p.177f.

[6] Curley (1978), p.181.

> ... I am now seeing light, hearing a noise, feeling heat. But I am asleep, so all this is false. Yet I certainly *seem* to see, to hear, and to be warmed. This cannot be false; what is called 'having a sensory perception' is strictly just this, and in this restricted sense of the term it is simply thinking. (CSM II 19)

This passage and Article 9 of Part I of the *Principles of Philosophy* (*"What is meant by 'thought'"*) (CSM I 195) are cited by Anthony Kenny as evidence of Descartes' commitment to the view that ideas "can be mistaken, but their occurrence and content cannot be doubted."[7] Margaret Wilson concurs, but observes that (as the passage from the *Principles* suggests) Descartes wants here not only "to make the claim that *cogitatio* judgments are absolutely certain and incorrigible," but also "to establish the thesis that *sensation* can be viewed as a type of thought ..."[8] Wilson's emphasis on "judgments" here is dubious, since Meditation III calls into doubt the reliability of judgment. If anything is regarded as known "incorrigibly," in these passages, it is the occurrence and content of "ideas," as Kenny suggests.

Still, one might wonder whether the content and the type-identity of sense-ideas--of seeing, hearing, being warmed, etc.--are meant to fall to the evil-demon hypothesis, given the confused and obscure character of which Descartes is concerned to convict the senses. (His criticisms of the senses usually seem aimed at their *veridicality*.) Indeed, one might question whether the self is supposed to be able to distinguish *any* idea-types; perhaps Descartes takes the self at this point in the *Meditations* to have certainty only of its thinking in general.

This last suggestion is too strong, since in addition to the *cogito* Descartes seems to hold that a number of "simple natures"[9] are intuitively certain, "manifest by the natural light" (CSM II 28), as he says of the scholastic causal principle in Meditation III. So, at least some particular ideas can be distinguished from each other. But, moreover, both Meditations II and III contain discussions that suggest

---

[7] Kenny (1987), p.71.

[8] Wilson (1978), p.75.

[9] *Rules for the Direction of the Mind*, Rule 12, CSM I 44f. The relation between the *cogito* and these simple natures known "*per se*" is debatable. Gueroult seems to suggest that (initially) even the latter fall to the deceiver-hypothesis. The *cogito* then serves as immediate ground for the other simple natures. On Gueroult's account the simple natures are not dubitable in themselves, in the sense that "by nature they cannot be doubtful" (Gueroult, 1984, p.18). But while they are not subject to natural doubt--while they are certain in themselves--there is no (metaphysical) certainty of this (natural) certainty, according to Gueroult. On this view, the *cogito* is both naturally and metaphysically certain, and it restores the certainty that metaphysical doubt removes from the simple natures.

In the same ballpark of Cartesian interpretation (both explore and emphasize the nature and importance of Cartesian *method*), Schouls does not take metaphysical doubt to extend to simple natures, which seem to enjoy the same certainty as the *cogito* on his reading. See Schouls (1980), p.109. For a revision of his reading see Schouls (1993).

the possibility of distinguishing types of ideas (and so, knowing when tokens of those types occur) and of knowing their contents with certainty. Such evidence appears in Meditation II in an earlier part of the paragraph in which we first met with "seeming" perceptions:

> Is it not one and the same 'I' who is now doubting almost everything, who nonetheless understands some things, who affirms that this one thing is true, denies everything else, desires to know more, is unwilling to be deceived, imagines many things even involuntarily, and is aware of many things which apparently come from the senses? Are not all these things just as true as the fact that I exist, even if I am asleep all the time, and even if he who created me is doing all he can to deceive me? (CSM II 19)

These questions are rhetorical, and Descartes quickly gives them affirmative answers, of which the discussion of seeming to see, to hear, etc. is a part. It is significant that their affirmation comes despite the operative status of the deceiver-hypothesis.

Further evidence is offered by Meditation III in which, after reintroducing the evil demon, Descartes does an inventory of ideas in order to determine "which of them can properly be said to be the bearers of truth and falsity" (CSM II 25). He proposes two distinct ways of categorizing thoughts: according to their intrinsic character and according to their apparent origins. It is the former that is relevant here. "Some of my thoughts," says Descartes, "are as it were the images of things ..." (CSM II 25). Others include "something more than the likeness" of the thing that is "the object of my thought," and some of these are called "volitions or emotions, while others are called judgments" (CSM II 26). That the contents of some of these ideas are known with certainty (at least, given the *cogito*) is suggested by part of the paragraph that follows:

> Now as far as ideas are concerned, provided they are considered solely in themselves and I do not refer them to anything else, they cannot strictly speaking be false; for whether it is a goat or a chimera that I am imagining, it is just as true that I imagine the former as the latter. As for the will and the emotions, here too one need not worry about falsity; for even if the things which I may desire are wicked or even non-existent, that does not make it any less true that I desire them. Thus the only remaining thoughts where I must be on my guard against making a mistake are judgments. (CSM II 26)

And we might also consider a long passage from the First Meditation in which, prior to the application of metaphysical doubt, ideas of sense are treated as composed of simple natures:

> [I]f perhaps they [painters] manage to think up something so new that nothing remotely similar has ever been seen before--something which is therefore completely fictitious and unreal--at least the colours used in the composition must be real. By similar reasoning, although these general kinds of things--eyes, head, hands and so on--could be imaginary, it must at least be admitted that certain other even simpler and

> more universal things are real. These are as it were the real colours from which we form all the images of things, whether true or false, that occur in our thought.
>
> This class appears to include corporeal nature in general, and its extension; the shape of extended things; the quantity, or size and number of these things; the place in which they may exist, the time through which they may endure, and so on. (CSM II 13f)

Notice that it is by "similar reasoning" that both colours and the subject-matter of "arithmetic, geometry and other subjects of this kind" (CSM II 14) are held to be "real" prior to the meddling of the evil demon. Ideas of sense, it would seem, carry a certainty with respect to their components that is of a piece with the certainty of simple mathematical truths, such as "two and three added together are five, and a square has no more than four sides" (CSM II 14). About the occurrence and content of such ideas there can be no natural doubt for an attentive mind because "these simple natures are all self-evident and never contain any falsity" (CSM I 45). And if colours and the like do not receive much consideration during the ensuing discussion, then this might be explained by the fact that Descartes will hold in Meditation VI that awareness of *these* simple natures is occasioned by the conjunction of the soul with a body.[10]

Now, the passages that I have offered are not conclusive evidence for Descartes' acceptance of (ii), but they are very suggestive of something like thesis (ii).[11] It should be emphasized, also, that (ii) includes reference to the need for attention if there is to be such incorrigibility. For only by "attentive and repeated meditation" (CSM

---

[10] See Gueroult (1984), p.16. See also Rule 12 (CSM I 44f.).

[11] Curley argues that such passages can be explained away by recognizing that Descartes accepts them only for the sake of argument against the sceptic. (See Curley, 1978, pp.170-193.) This claim is connected--I take it--with Curley's view that Descartes' writings should be understood against the background of a resurgent interest in scepticism during the seventeenth century. (See *ibid.*, pp.9-20.) I do not wish to take issue with Curley's broader project, but it does seem to me that he under-emphasizes the importance of "sceptical" doubt as a methodological tool for Descartes, especially in the *Meditations*, where it is said to have the great benefit of "freeing us from all our preconceived opinions, and providing the easiest route by which the mind may be led away from the senses" (CSM II 9). For similar concerns see Reynolds (1992). (Reynolds says in passing that Davidson and Putnam have "Cartesian starting point[s]" (*ibid.*, p.188n6), a comment that does not recognize the profoundly anti-Cartesian nature of their positions.) It is by testing beliefs and ideas for their dubitability that Descartes carries out his project of *analysis*--certainty and simplicity are closely linked for Descartes. For discussions of doubt and analysis see Gueroult (1984) and Schouls (1980).

II 43) can I avoid error. So, (ii) differs from the less plausible claim which, Curley suggests, is sometimes attributed to Descartes.[12]

Why is thesis (ii) of interest? We saw above that Descartes' account of the self as a thinking substance amounts to treating the self as having the same sort of objectivity that the metaphysical realist gives to the so-called "external" world--with the important difference that the nature and existence of the self are not independent of its *own* epistemic capacities. Now, if (i) represents a commitment to the metaphysical *objectivity* of the self, (ii) is a thesis about its *subjectivity*. It is a claim that the self does not exist apart from its ability to know itself with certainty. But the combination of (i) and (ii) is an extraordinary philosophical position. The metaphysical realist about the external world does not typically suppose (unless she is a traditional theist) that any particular agent has infallible epistemic access to the world. The nature and existence of the world are treated as independent of the abilities of all to know and describe them--we are vouchsafed no certainty in our dealings with the "external" world. Why the tremendous difference between the two cases?

In part Descartes recognizes what at some level must be true: there is a presumption that I know my own mind better than others do, and better than I do others'. My claims about my intentional attitudes carry at least a *prima facie* authority that neither others' claims about me, nor my claims about others do. But, as I shall argue in Chapter 7, one need not embrace any thesis of incorrigibility or infallibility to account for this "first-person authority." Another component in the explanation of Descartes' commitment to (i) and (ii) is his demand for *certainty*, a demand for which he has been often criticized by philosophers in the twentieth century. However, the perceived need for certainty is itself something that makes sense only given Descartes' commitment to metaphysical realism and the objective-subjective dichotomy. For unless sceptical doubt as a methodological tool for analysis can get off the ground, certainty of the sort that Descartes has in mind can never appear as more than a tempting mirage. Only if the subjective is taken to stand "over against" the objective, does it make sense to think that the subject has some essentially distinct and special access to itself that it does not have to the object. Of course, on my view there are some things of which we can be certain, in the sense that they are not subject to coherent doubt--"I have not always been a brain in a vat," for example. But this is not the certainty of an "exact coincidence between my thought and existence" in which "the object posited is nothing but the subject."[13] It is not a certainty that looks on evil demons and is never shaken, because it rules out the coherence of evil demons from the start.[14]

---

[12] *I.e.*: "Our beliefs about our mental states are incorrigible, in the sense that, necessarily, if a person believes that he is in a particular mental state, he is in that state" (Curley, 1978, p.171).

[13] Gueroult (1984), p.27.

[14] The kind of "certainty" admitted by my account is similar in some ways to Descartes' "metaphysical" certainty--at least, to the extent that the latter is captured by the self-defeating nature of some kinds of doubt. There is, indeed, a sense in which I cannot coherently doubt my own existence, as Descartes held, but it is a

II. Hume: The Illusions of Subjectivity

One reaction to Descartes' conception of the self and our knowledge of it is Hume's. But his also presents us with a peculiar mixture of the objective and the subjective. If I have knowledge of my self, and if my self is always present, an enduring, unchanging substance that underlies my thinking and perceiving, Hume maintains, then there must be a persistent representation of this self, for knowledge is had only by means of "ideas" and "impressions." But this raises a difficulty:

> [S]elf or person is not any one impression, but that to which our several impressions and ideas are suppos'd to have a reference. If any impression gives rise to the idea of the self, that impression must continue invariably the same, thro' the whole course of our lives; since self is suppos'd to exist after that manner. But there is no impression constant and invariable. (THN 251)

There is no persistent idea of the simple substantial self,[15] says Hume, and hence no knowledge of the self. He says in the *Treatise*,

> For my part, when I enter most intimately into what I call *myself*, I always stumble on some particular perception or other, of heat or cold, light or shade, love or hatred, pain or pleasure. I never can catch *myself* at any time without a perception, and never can observe anything but the perception. (THN 252)

---

[14] sense that prevents my knowledge of others and of the rest of the world from being subject to coherent doubt, as well. This is an attempt to make coherent the insight of the *cogito*. Such coherence, however, is purchased without recourse to theism and at the cost of abandoning metaphysical realism. "Certainty" in this sense amounts to being in an epistemic position with regard to some proposition, such that nothing could count as a good reason for doubting that proposition. (Any reason I could have to doubt the proposition in question is a reason for doubting my doubt.)

The "psychological" or "natural" certainty that Descartes regards the "simple natures" to possess, while needing "metaphysical certainty" to escape the deceptive mischief of the evil demon, has no role to play in my account. Such certainty amounts for Descartes, not to immunity from coherent doubt, but to a kind of immunity from error: I cannot be mistaken about the occurrence (and content, I have said) of my current ideas. Moreover, in its applicability to "ideas" this sort of certainty is supposed to hold for a great many cases, which, taken individually, do not escape coherent doubt in my account. The extension of such certainty beyond the occurrence and content of ideas depends for Descartes on the further premise of God's perfection.

[15] The Cartesian response, on Gueroult's reading, would be to argue for God's existence and appeal to divine benevolence to make up for the absence of a persistent idea of self. See Gueroult (1984), pp.103ff. I doubt Hume would have found Meditation III compelling.

Hume, I suggest, draws out the sceptical conclusion that is a tacit threat in Descartes' treatment of the self as a substance. The idea that the self is *a metaphysically objective thing* makes the idea of privileged access to it mysterious.

Now, it might be objected that Hume's scepticism about the self comes solely from his representationalism, and that we should think of self-knowledge (of the sort that would reveal the self as substance) as some kind of immediate relation of the self to itself. But if Hume's modest claim that he can find no impression of the self as persistent and unchanging is compelling, it is, I suspect, because there is no immediate awareness of the self as substance either--no awareness that Hume would have been tempted to drag into the category of "impressions."[16] Even if we grant the *cogito* to embody an immediate awareness of self, two problems remain. First, as Descartes worried, "I am, I exist--that is certain. But for how long? For as long as I am thinking" (CSM II 18). The problem with the certainty given in the *cogito* is that it is fleeting. "[O]nce my mind stops fixing on the Cogito in order to fix on something else, the point of support disappears in the darkness of universal doubt, carrying with it the whole chain of reasons."[17] Secondly, it is a further, questionable step to conclude that the self is a *substance*. It was the unjustified move--mediated or not--from the one claim to the other that Kant criticized as the First Paralogism of Pure Reason:

> That the 'I', the 'I' that thinks, can be regarded always as *subject*, and as something which does not belong to thought as a mere predicate, must be granted. It is an apodeictic and indeed *identical* proposition: but it does not mean that I, as *object*, am for myself a *self-subsistent* being or *substance*. (B 407)

Despite Hume's scepticism of the self--based, I have suggested, on the idea that self-knowledge must be knowledge of a metaphysically objective thing--he is by no means free of epistemically neurotic combinations of the objective and the subjective. First, impressions and ideas are for Hume "distinct existences" (THN 636); indeed, criticizing the Cartesian doctrine of substance, he ventures to say that ideas and impressions satisfy Descartes' account of substance:

> For thus I reason. Whatever is clearly conceiv'd may exist; and whatever is clearly conceiv'd after any manner, may exist after the same manner. ... Again, every thing, which is different is distinguishable, and every thing which is distinguishable, is separable by the imagination. ... My conclusion ... is, that since all our perceptions are different from each other, and from every thing else in the universe, they are also distinct and separable, and may be consider'd as separately existent, and may exist separately,

---

[16] There is, importantly, awareness of oneself as a biological organism, but neither Hume nor Descartes accepts any such appeal, since my body is one more object in the empirical world, and I can be no less mistaken about its existence than about the rest of the world.

[17] *Ibid.*, p.105.

> and have no need of any thing else to support their existence. They are, therefore, substances, as far as this definition explains a substance. (THN 233)

Whether we should take Hume to subscribe to this view or merely to regard it as an immanent critique of Cartesian assumptions is an interpretive issue that I shall not explore here. But there is, at least, the following parallel between such reasoning and Hume's clearer commitments on the nature of perceptions. My perceptions are private, and they exist independently of the epistemic capacities of others; furthermore, I cannot be mistaken about my perceptions.

'Privacy' is not a part of Hume's philosophical jargon, of course, but 'immediacy' and 'certainty' are:

> The only existences of which we are certain, are perceptions, which being immediately present to us by consciousness, command our strongest assent, and are the first foundations of all our conclusions. (THN 212)

And Hume indicates that it is only by means of perceptions that we can acquire beliefs about objects. "[N]o external object can make itself known to the mind immediately, and without the interposition of an image or perception" (THN 239), he writes. "[N]othing is present to me but particular perceptions ..." (THN 634). Perhaps his most explicit avowal that we cannot be mistaken about our ideas and impressions--however much we may be misled about the "external" world--comes in Section II, of Part IV of Book I of the *Treatise*. There he says that

> ... all sensations are felt by the mind, such as they really are, and ... when we doubt, whether they present themselves as distinct objects, or as mere impressions, the difficulty is not concerning their nature, but concerning their relations and situation. (THN 189)

He continues, retracting the last clause of the preceding passage:

> [T]hey appear, all of them, in their true colours, as impressions or perceptions. And indeed, if we consider the matter aright, 'tis scarce possible it shou'd be otherwise, nor is it conceivable that our senses shou'd be more capable of deceiving us in the situation and relations, than in the nature of our impressions. For since all actions and sensations of the mind are known to us by consciousness, they must necessarily appear in every particular what they are, and be what they appear. Every thing that enters the mind, being in *reality* a perception, 'tis impossible any thing shou'd to *feeling* appear different. This were to suppose, that even where we are most intimately conscious, we might be mistaken. (THN 190)[18]

---

[18] Penelhum notes, however, that Hume seems less confident of this degree of privileged access in his discussion of calm passions in Book II and in his account of confused impressions of hatred and moral disapproval in Book III. See Penelhum (1975), p.202n3.

"Knowledge" of our perceptions is, thus, a purely subjective affair, provided the nature and occurrence of my perceptions are independent of the epistemic capacities or the beliefs of others. For where I cannot possibly be mistaken, there is no distinction between my thinking that I am correct (about the nature of my impressions) and my being correct. If I think that I am right where I am "most intimately conscious," then I *am* right. There is no objectivity regarding my perceptions, because objectivity presupposes independence from my (indeed, anyone's) beliefs--*i.e.*, it presupposes the chance of error.

That my perceptions are independent of the abilities of others to know or describe them seems to follow from a number of Hume's claims. For instance, he sometimes seems to view individual perceptions as capable of independent existence, much as is suggested in his critique (above) of Cartesian substance. "[W]e may observe," he argues,

> that what we call a *mind*, is nothing but a heap or collection of different perceptions, united together by certain relations, and suppos'd, tho' falsely, to be endow'd with a perfect simplicity and identity. Now as every perception is distinguishable from another, and may be consider'd as separately existent; it evidently follows, that there is no absurdity in separating any particular perception from the mind; that is, in breaking off all its relations, with that connected mass of perceptions, which constitute a thinking being. (THN 207)

And in the Appendix we find reasoning similar to that which he employs in showing that perceptions count as "substances":

> Whatever is distinct, is distinguishable; and what is distinguishable, is separable by the thought or imagination. All perceptions are distinct. They are, therefore, distinguishable, and separable, and may be conceiv'd as separately existent, and may exist separably, without any contradiction or absurdity. (THN 634)

Elsewhere Hume seems more hesitant about this possibility. Indeed, he takes it as a contributing psychological factor to the doctrine of "double existence" that perceptions can be shown to be dependent on our senses for their existence:

> When we press one eye with a finger, we immediately perceive all the objects to become double, and one half of them to be remov'd from their common and natural position. But as we do not attribute a continu'd existence to both these perceptions, and as they are both of the same nature, we clearly perceive, that all our perceptions are dependent on our organs, and the disposition of our nerves and animal spirits. (THN 211)

However, since perceptions are "internal and perishing existences" (THN 194), the possibility that mine might linger for another to perceive seems ruled out. Consequently, whatever Hume's view on the "substantiality" of perceptions, my impressions and ideas are such that everyone else could be completely mistaken about them. Hume's position is a paradoxical one of subjectivity without self-knowledge.

III. Kant: Self-Knowledge Is Not Self-Sufficient

Kant moves away from the problems incurred by a dichotomous treatment of the subjective and the objective, such as we find in Hume and Descartes. But he does not abandon that dichotomy outright. One hint of this lies in his meta-ethical views, where a metaphysical notion of the self seems important to his talk of moral agents as noumenal beings in a kingdom of ends. Less obvious, but no less significant, is his commitment to a sharp distinction between formal conceptual schemes and empirical content, or--better--between *one* exclusive conceptual scheme and its empirical content. As Davidson has observed, the scheme-content and objective-subjective dualisms "have a common origin: a concept of mind with its private states and objects."[19] Kant's critique of the latter dichotomy is not decisive, for he remains committed to the dualism of scheme and content. Only by insisting that we *must* use certain forms of judgment, certain pure concepts of understanding, does he avoid the troubles of alternative, incommensurable conceptual schemes. Once this artifice is abandoned, the dialectic of realism and relativism reasserts itself. Thirdly, Kant's insistence on one sharply bounded conceptual scheme is encouraged by another relic of the Cartesian tradition--the 'idea'-idea. The fact that he still takes the mind to do its business with "representations" (*Vorstellungen*), not with language, makes it easier to avoid a simple truth: we do not all use the same language. But his critique of Descartes and Hume remains insightful. Let us consider it.

Kant goes a step beyond Hume.[20] Not only does he think that there is no persistent idea or impression--no ever-present representation--of the self, but he challenges the objective-subjective dichotomy. The self, he says, is not a substance, but neither is it simply "a bundle or collection of different perceptions" (THN 252). What emerges is not a more radical scepticism than Hume's, nor a Berkeleian plea that we have a "notion" of self,[21] rather than an idea. Kant conceives of the self as a thing that is known "only as it appears to itself" (B 69), not as it is in itself, and also "as it is affected by itself" (B 69).

Consider Hume's claim that when he looks for an impression of self, all that he finds is "some particular perception or other, of heat or cold, light or shade, love or hatred, pain or pleasure." In the Preface to the *Critique's* 1787-edition Kant holds a similar thesis.[22] The "whole material of knowledge, even for our inner sense," he writes, is derived from "things outside us" (B xxxixa). And later he tells us that "the representations of the *outer senses* constitute the proper material with which we occupy our mind" (B 67).

Outer sense, for Kant, is our mode of intuiting the world of spatio-temporal objects, but our contact with that world is not to be thought of on the model of "transcendental realism" (A 369). We know objects

---

[19] Davidson (1989), p.163.

[20] In what follows my thinking has been greatly influenced by Allison (1983). I draw here on Hymers (1992).

[21] Berkeley (1982), Part I §§27,140,142.

[22] He excludes "feeling of pleasure and pain, and the will" (B 66) on the ground that they are "not ... knowledge" (B 66).

not as they are in themselves, but only as they appear to us: "We know nothing but our mode of perceiving them" (A 42/B 59).[23]

Outer sense is distinct from *inner sense* "by means of which the mind intuits itself or its inner state" (A 22/B 37). But like outer sense, inner sense gives me knowledge, not of things in themselves, but of things as they appear to me. More correctly, inner sense provides me with what knowledge of *myself* is possible, and so, it provides me with knowledge of myself *as I appear to myself*, not of myself as a thing in itself. But the waters grow muddier, for Kant says that the mind

> ... intuits itself not as it would represent itself if immediately self-active, but as it is affected by itself, and therefore as it appears to itself, not as it is. (B 69)[24]

There is more here than the ready parallel with outer sense and our knowledge of "the world" as construed by transcendental idealism. For, we are told, the mind intuits itself "as it is affected by itself". Let us see if this can be clarified at all.

I mentioned above that Kant shares Hume's rejection of the suggestion that I have a persistent and distinct representation of myself, an "impression constant and invariable". Consider, now, this revealing passage from the first edition of the *Critique*:

> For in what we entitle 'soul', everything is in continual flux and there is nothing abiding (*nichts Bleibendes*) except (if we must so express ourselves) the 'I', which is simple solely because its representation has no content, and therefore no manifold, and for this reason seems to represent, or (to use a more correct term) denote, a simple object. (A 381-382)

Some of these remarks need emphasis and explication. First, the 'I' to which Kant refers is a representation without a *manifold*, without *content*. Consequently, it *seems* to denote (*zu bezeichnen*) a "simple object;" however, as I shall explain, it does not *actually* denote such an object. Self-knowledge, according to Kant, is made possible by inner sense, but "Inner sense ... yields indeed no intuition of the soul itself as an object ..." (A 22/B 37). On the contrary, knowledge of the self requires the intuition of a manifold of representations: "for knowledge of myself I require, besides the consciousness, that is, besides the thought of myself, an intuition of the manifold in me" (B 158). Moreover, Kant holds that inner sense depends for its manifold on outer sense whose representations "constitute the proper material with which we occupy the mind" (B 67); there is no direct intuition of the self, merely an intuition of one's states, and the only material for knowledge here seems to be one's representations of outer things, including one's body and its relations to other things.

Hence, a "representation" without content, without a manifold, could provide me with no knowledge of myself. And this is why it could only

---

[23] This transcendental idealism is not the "pragmatic realism" I have been advocating, since Kant retains the regulative idea of transcendental reality about which nothing can be said or thought.

[24] See also B 153-159.

seem to represent a simple object--*viz.*, a Cartesian self. But its very emptiness is simple in a way which, Kant thinks, leads us to mistake the 'I' for a representation of a Cartesian *ego*, a mental substance that could exist apart from any embodiment. This, as we saw earlier, is the error of the First Paralogism of Pure Reason (A 366-380/B 409). But what is this 'I' that it should provoke such an error?

The 'I' to which Kant refers here is what he calls elsewhere in the *Critique* the "transcendental unity of apperception" (A 108,118).[25] It is "a merely *intellectual* representation of the spontaneity of a thinking subject" (B 278) in which "I am conscious of myself, not as I appear to myself, nor as I am in myself, but only that I am" (B 157). There is for Kant a stark divide between form and content, between intellect and intuition, the latter of which can be only empirical. Whereas some of his predecessors accepted the idea of an intellectual intuition of the self and of *a priori* truths, Kant insisted that human knowledge required empirical intuition, though acquaintance with the empirical would always be structured by the pure concepts of understanding, conditioned in turn by transcendental apperception.

I think that there is a fairly straightforward way of grasping what this "original apperception" is supposed to be. Kant speaks of it as "an objective condition of all knowledge" (B 138), but also indicates that it is "an identical, and therefore analytic, proposition ..." (B 135). This is so, he claims, because "it says no more than that all *my* representations in any given intuition must be subject to that condition under which alone I can ascribe them to the identical self as *my* representations" (B 138). Kant's point is deceptively simply. To be able to say correctly that several thoughts are *mine*, I must be able to be conscious of them--or better--I must be capable of thinking of these thoughts as united together in a single consciousness, as being *all* mine. It would be insufficient for me to be able to be conscious of each thought as I was having it, but conscious of having no other. That would just be an awkward way of saying that I had that thought. Rather, I must be able to tell one from the other and to be conscious of having these distinct thoughts as my own. But in binding together these thoughts in a "synthetic unity", I "bind" together myself, enabling an awareness of myself as distinct from other things.

However, this 'unity' of consciousness is not very substantial--indeed, it is not *substantial* at all in Kant's view. It is merely a formal condition of the possibility of thinking of objects and of having empirical knowledge. Kant maintains that Descartes goes astray in wanting to move from the *cogito* to conclusions about himself--that he is a thinking *thing*, a substance that can exist apart from matter. The "'I think'" (B 131) or the "'I am'" (B 277), which Kant equates with pure apperception, is a consciousness of self at its barest, a consciousness that becomes actual only through appearances--*i.e.*, only with the aid of sensory input. Still, Kant takes it to remain distinct

---

[25] Kant's terminology varies considerably, embracing variations such as "*transcendental apperception*" (A 107), "transcendental ground of the unity of consciousness" (A 106), "formal unity of consciousness" (A 105), "necessary unity of consciousness" (A 109). "pure apperception" (A 116-117/B 132), "original synthetic unity of apperception" (B 131), "original apperception" (B 132), "transcendental unity of self-consciousness" (B 132), "necessary unity of apperception" (B 135), "objective unity of self-consciousness" (B 139), etc.

from the *empirical* unity of apperception,[26] which lies in a specific consciousness united in specific representations of empirical objects, not in the general condition that makes this specific unity possible.

Now, it must be emphasized that this constraint on the possibility of experience is no return to the metaphysical enigma of Cartesian substance. Kant explicitly contrasts original apperception with the notion of thinking substance--the lesson of the First Paralogism:

> The analysis, then, of the consciousness of myself in thought in general, yields nothing whatsoever towards the knowledge of myself as object. The logical exposition of thought in general has been mistaken for a metaphysical determination of the object. (B 409)

Kant does not advocate a metaphysical unity, *i.e.*, the unity of a thing that can exist, complete to itself, apart from anything else (except, maybe, a benevolent deity). Descartes, he seems to suggest, confuses what we might now call the "grammatical subject" of my judgments, 'I', with a metaphysical category: *substance*.[27]

It is not just the objective pole of Descartes' conception of the self from which Kant distances himself. The subjective pole, embodied in what he takes to be Descartes' commitment to immediate self-awareness, is put to skilful critique in Kant's "*chef d'oeuvre* of compressed obscurity,"[28] the Refutation of Idealism. The position advanced there is a striking anticipation of the reasoning which, I said in Chapter 4, is present in Wittgenstein's "private language argument."

The peculiarly Kantian insight of the Refutation is that self-knowledge *presupposes* temporal determination of the self, so that one cannot appeal to one's self-knowledge for time-determination. Kant holds, remember, that I know myself as I appear to myself and as I am affected by myself. We can now say more clearly what is meant by this. Kant insists that the possibility of empirical knowledge implies the ability to join distinct representations in a single consciousness. But, he says, self-knowledge consists of knowledge of one's own representations. "[O]nly in so far as I can grasp the manifold of the representations in one consciousness, do I call them one and all *mine*" (B 134). If I could not bind my disparate representations together in a single consciousness "... I should have as many-coloured and diverse a self as I have representations of which I am conscious to myself" (B 134). By conjoining my representations of the empirical world in a more complex representation I tacitly assign myself a position from which the world is understood, and when I try to reflect explicitly on myself, I combine in a single consciousness some of my various encounters with the world, as I seem to remember them. I must, then, "affect" myself in order to "intuit" myself, because forming a complex representation of my past representations *adds* to my representations, both in quantity (there are now more representations) and in quality (those representations are now regarded differently). The idea here--

---

[26] See A 107,115/B 132,140.

[27] See B 407.

[28] Bennett (1966), p.202.

that self-knowledge "does not leave its object unchanged,"[29] in Charles Taylor's words--is one to which I shall return in Chapter 8.

This self-affecting is achieved by combining my representations of empirical objects and events into a more complex representation of myself as the holder of these representations, as the self whose body is among those empirical objects represented. But in so representing myself, I assign a determinate temporal ordering to the events or states represented. How, Kant wants to know, do I manage to do this? His question is directed at Cartesianism. In particular, he seeks to undermine Cartesian scepticism--"problematic idealism" (B 274)--by beginning with a premise to which he thinks the Cartesian must accede: "I am conscious of my own existence as determined in time" (B 275).

Cartesian scepticism, recall, is best understood as the proposal of an alternative explanatory hypothesis for the nature and variety of our experience; it is presented as a *real possibility* that is meant to explain some phenomenon of which we are assumed to have knowledge. Since I have no permanent impression of self to serve as a temporal standard against which I can order my representations, and since I cannot perceive time itself, I am thrown back on my representations for an answer. But taken formally, as states of my consciousness, my representations offer no unequivocal standard, for I can order my representations by reference to those same representations only in a trivial manner. *Any* ordering will be trivially justified, if the only criterion of justification is that it conform with itself. Nor is it enough to say that my representations are given their ordering simply in virtue of their being uniformly caused. One can imagine this proposal as an explanatory hypothesis, and the Cartesian would happily accept it, since the deception practiced on me by an evil demon might be extremely regular in its character. But the mere occurrence of my representations in a regular order does not show that my *consciousness of them* is so ordered. That requires a further complex representation, without which it follows only that I *have* certain representations, not that I am conscious of myself as I am determined in time.

If my representations are to be ordered by reference to themselves, then it must be by reference to themselves taken materially--*i.e.*, by reference to their "content." "All determination of time presupposes something *permanent* in perception" (B 275), says Kant. But the only thing permanent--or relatively abiding--in perception is the empirical realm of spatio-temporal objects, and it is relative to this that I can order my representations in a more complex representation. My representations, *qua* states of myself, can be ordered by correlating them with successive states of empirical objects.

Let us suppose that, in fact, there is no independent normative constraint on the ordering of my representations--that there is no correct order over and above the one that I happen to establish. If this is the case, then there is no distinction between thinking I am right about the order of my representations and being right. I cannot be mistaken. But then, I can from one moment to the next change my mind about what constitutes a correct ordering. Or rather, what we would normally regard as a change of mind, as a new ordering, I can regard as consistent with my first ordering, and there are no grounds for criticizing me. At t_1 I may suppose that several of my conscious states, S_a, S_b, S_c, belong in the order, S_a, S_b, S_c, and I will be

---

[29] Taylor (1985), p.36.

correct. Then at t_2 I may decide that they belong in the order, S_c, S_a, S_b, and I will be correct *again*. If it is objected that S_c, S_a, S_b, is an *incorrect* ordering, because it does not conform with S_a, S_b, S_c, then a distinction between thinking I am right and being right has been reintroduced--a standard independent of my complex representation of the order of my conscious states is being appealed to. I am well within my rights to say that my second judgment *accords perfectly* with my first. There is no distinction here between my being right and my thinking I am right, because there is no standard beyond that which I establish. But if this is the case, then *anything at all* can be taken to accord with my first judgment concerning the order of my conscious states. Hence, there is no fact of the matter concerning that order.

Kant's Refutation is intended as an anti-sceptical argument: it is a necessary condition of self-knowledge that I know about an objective realm of spatio-temporal things. As Kant says, "inner experience is itself possible only mediately, and only through outer objects" (B 277). But notice that the possibility that I can be wrong about myself has been introduced as a crucial step in the argument. The subjective pole of Descartes' account of the self has also been rejected.

Since Hume is also committed to the view that I cannot be mistaken about the nature and occurrence of my representations--not to mention their "situation and relations" (THN 190)--it is worth asking what consequences, if any, the Refutation has for Hume's views. It seems clear that Kant's first premise is a non-starter for Hume, who would deny that "I am conscious of my own existence as determined in time" (B 275). But we can replace this premise with one that coheres with Humean epistemology to produce a variation on Kant's argument, thus:

> 1. I am conscious of my representations as determined in time.
> 2. "All determination in time presupposes something *permanent* in perception" (B 275).
> 3. "But this permanent cannot be an intuition in me. For all grounds of determination of my existence which are to be met with in me are representations; and as representations themselves require a permanent distinct from them, in relation to which their change ... may be determined" (B xxxixa).
> 4. "Thus perception of this permanent is possible only through a *thing* outside me and not through the mere *representation* of a thing outside me ..." (B 275).
> ---
> 5. Therefore, objects exist in space outside me.

The acceptability of premise 1 follows from Hume's remark that it is not "conceivable that our senses shou'd be more capable of deceiving us in the *situation and relations* [my emphasis], than in the nature of our impressions" (THN 190). Indeed, according to Hume, we derive the abstract idea of time from the succession of perceptions:

> As 'tis from the disposition of visible and tangible objects we receive the idea of space, so from the succession of ideas and impressions we form the idea of time ... (THN 35)

So, there can be no objection to our new first premise, if we take it to record merely an order of succession. Premise 3 is also acceptable

to Hume (though perhaps for non-Kantian reasons), for he agrees with Kant that there is no "impression constant and invariable" (THN 251) and, hence, "no permanent intuition" (B 292). The status of premise 4 would seem to turn on the acceptability of premise 2; so, the issue is whether or not Hume would accept 2. This is a problem.

Premise 2 is in intermediate step in Kant's argument in the First Analogy of Experience (A 182-189/B 224-232). His task there is to show that our thinking according to the category of substance is necessary for the possibility of empirical knowledge, and his strategy is to show that temporal relations presuppose substance. What matters most here is the claim that we can have knowledge of temporal succession only by reference to a relatively enduring temporal standard. As Kant observes, I cannot perceive time directly--I have no impression of time. With this much Hume agrees: "[N]or is it possible for time alone ever to make its appearance, or be taken notice of by the mind" (THN 35). However, this raises the question of whether something persistent *in* perception is needed, or whether merely a succession of distinct impressions suffices for an idea of succession.

I said that Kant sees Descartes as confusing the grammatical subject of my judgments with the category of substance. But if Descartes hypostasizes the subject, Hume underestimates its import. Kant and Hume agree that "in bare succession existence is always vanishing and recommencing" (A 183/B 226)--perceptions are "perishing existences" (THN 194) which "succeed each other with an inconceivable rapidity, and are in a perpetual flux and movement" (THN 252). What I think Kant would regard as the problem for Hume lies in the need to *compare* and contrast successive impressions if they are to be justifiably regarded as successive. Hume himself seems to say that the relevant succession is one of qualitatively different impressions. Time, he says, "is always discover'd by some *perceivable* succession of changeable objects" (THN 35). But if I am to compare an impression at t_2 with an impression at t_1 I must, as it were, "hang on" to the impression from t_1 and "hold it up beside" the impression from t_2. Only then can I have justification for thinking that they do or do not differ. The moment at which two billiard balls collide ought not to be thought of as following the moment at which one "approaches" the other, unless the representation of the "approaching" ball can be thought together in a series with that of the two balls in contact.

But this comparison of successive impressions is just what Kant intends by the ability to "unite a manifold of given representations in *one consciousness*" (B 133). Knowledge of succession presupposes the transcendental unity of apperception. This unity, as we have seen, is not an actual unity; it is the formal condition of the possibility of an actual or empirical unity of apperception. But such an empirical unity is what is embodied in any *particular* comparison of successive representations--my representation of the colliding billiard-balls, for example. And this matters, because once this empirical unity has been granted, Kant has the ingredients he needs for the final step in his critique of Hume. Once the unity of apperception has found its way into the argument, there is no obvious reason why the synthesis that unites successive impressions to yield the idea of their succession should not be applied again to those individual empirical unities of apperception to give a more general unity--a unity which amounts for Kant to knowledge of the empirical self, as it appears to, and is affected by, itself in this synthesis. From here the Refutation runs

as before, and its tacit critique of consciousness applies, not just to a Cartesian conception of the self, but to a Humean one, as well.

I noted above that one of the consequences of Kant's Refutation is that "inner experience is itself possible only mediately, and through outer objects" (B 277). As well, on Kant's view it is possible to be mistaken about oneself. Self-knowledge claims are not incorrigible. But how mediated is inner experience, and to what degree are mistakes about myself possible? To answer these questions I shall turn from this brief historical survey to more recent work. And in the process my inquiries will take a turn that Kant's did not, a linguistic turn.

Chapter 7: The Objective and the Subjective in First-Person Authority

I have examined some modern roots of the problems that we meet when we try to apply a dichotomous understanding of the objective and the subjective to knowledge of the self. In this chapter I shall show that these odd mixtures of objectivity and subjectivity persist in current discussions of self-knowledge. As in earlier chapters, the therapy for the epistemic neurosis that arises from these blends begins by leaving the mistaken account of objectivity that is its cause. I shall focus on some recent work by Tyler Burge, arguing that it, too, attempts to combine objectivity and subjectivity without first dismantling the logical barriers that have been erected between them. I shall then turn again to the ideas of Donald Davidson. His account of first-person authority opens a way of circumventing the subjective-objective dichotomy. The capacity for self-knowledge turns out to be the capacity to give utterance to one's intentional attitudes. Self-knowledge depends on one's linguistic capacities, and since one can make linguistic errors, one can make errors about oneself, but not all or most of the time. This fact seems to place one in a relation with oneself that differs from one's relation with others. But I shall argue that this asymmetry between self-knowledge and knowledge of others is not absolute, but a matter of degree. There are cases in which it would be almost as odd to suppose that I were mistaken about the words of others as that I were mistaken about myself. People show degrees of "transparency" in the expression of their intentional attitudes, and the asymmetry between first- and other-person authority can be eroded or even eliminated in some cases. I shall argue that the degree of asymmetry varies as the inverse of the amount of background-knowledge and experience shared by the persons in question, and also as the inverse of the degree of justifiable trust among them. *Ceteris paribus*, if two people share a high degree of background-experience and knowledge, then their knowledge of each other will be more like their knowledge of themselves. Less interpretation is required in such cases than Davidson suggests. He is led to think that interpretation must always go on when I try to understand another by overestimating the importance of radical translation. Similarly, other things being equal, if two people stand in a relation of justifiable trust with each other, then their intentional phenomena will be more readily accessible to each other than if their trust is unjustifiable. The need for interpretation in other-knowledge again decreases. This claim will require a brief examination of the idea of "justifiable trust."

I. Anti-Individualism: Putnam and Burge

In Chapter 5 I speculated on sources of the politically motivated relativist's conflation of realism with scientism, suggesting that two historical cases of the contingent alignment of these doctrines might prompt the relativist's elision: on the one hand, some recent analytic philosophy, and on the other, some influential trends in Marxism and critical theory. In the latter case, I suggested, the objectivist side of the objective-subjective dichotomy infiltrates the understanding of the self. Accounts of ideology and false consciousness that portray the oppressed as utterly deluded in a way that serves the interests of their oppressors place the individual in a position where her only epistemic option seems to be to adopt a sceptical position and take it

as a real possibility that she knows nothing about the world, or at least the "social" world around her. Moreover, such views may seem to treat the individual as completely self-deluded, ignorant of her true needs and interests, and of the real part she plays in her society.

There is a parallel to this objectivist tendency in recent analytic philosophy.[1] It derives, perhaps, from the behaviourism popular earlier in this century, and it is displayed generally by theories of mind that hold it possible for a person to be entirely mistaken about the meanings of her own words. The trend has caught the attention of Tyler Burge and Donald Davidson, both of whom have tried to account for "first-person authority" without embracing the subjectivism of self-knowledge, popular in the Cartesian tradition. "The threat" to first-person authority, maintains Davidson,

> was there in Russell's idea of propositions that could be known to be true even though they contained 'ingredients' with which the mind of the knower was not acquainted; and as the study of the *de re* attitudes evolved the peril grew more acute.[2]

It is these *de re* attitudes that interest Burge. He claims not only to abandon "the traditional rationalist assumption" of subjectivism about self-knowledge, but to combine his less stridently subjectivist account with the view that "mental states" acquire their significance from the environment of the thinker who has them. His view is influenced by the theory of meaning that Putnam championed before falling from grace with metaphysical realism.

"[I]t was Hilary Putnam," says Davidson, "who pulled the plug."[3] In "The Meaning of 'Meaning'" Putnam argued that linguistic meaning is made up of a number of different elements, including extension. *Reference*, he claimed, is part of the meaning of a word, and it is fixed by factors beyond "the psychological state of the individual

---

[1] If Rorty is right, the parallel is a strong one, indeed. He suggests that Putnam's theory of direct reference, part of the theory of meaning I shall discuss below, was motivated by his now abandoned Marxist views. He writes in a review of Terry Eagleton's *Ideology*:

> Our way of thinking--that common to wet pragmatist liberals like me and dry postmodernist radicals like the Foucauldians --overlaps with a prominent historicist and pragmatist strain in Marx and Marxism. (Think of the eleventh thesis on Feuerbach, of the idea that morality is relative to class interests, and so on.) But it has been stoutly resisted by many Marxists--from Engels, through Lenin's criticisms of Berkeley and Mach in *Materialism and Empirio-Criticism*, to Milton Fisk's *Nature and Necessity* (a pre-Kripkean repudiation of Quinean holism and revivification of Aristotelian essentialism) and Hilary Putnam's politically-driven quest (in his Marxist period) for a physicalist theory of reference. (Rorty, 1992b, p.40)

[2] Davidson (1987), p.443.

[3] *Ibid.*

speaker."[4] Thus, a particular speaker might be wrong about the extension of, *e.g.*, 'water', but still be able to use the word.

Why should we think that reference is relevant to *meaning*? One reason arises from an account of meaning as *use* and its interactive conception of reference. That an expression or a word refers to a particular thing or sort of thing is a consequence of the roles that the word and the thing or sort of thing play in our lives. Reference is determined by practice, but in a sense practice is determined by reference, as well. The two go hand in hand--or rather, there is an internal connection between reference and practice, for without *referents* there is no practice. That lesson derives from the real impossibility that the Cartesian sceptic is right.[5] But this sort of reason does not appear in "The Meaning of 'Meaning'."[6]

Putnam's theory was motivated in part by a commitment to *scientific realism*: roughly the view that the entities postulated by scientific theories do in fact exist. Statements about electrons are true (or false) in virtue of the fact that they correspond (or fail to) with the world. The metaphysical version of scientific realism (scientific realism plus metaphysical realism) takes the world to have its existence and nature apart from our epistemic capacities and views the correspondence in question as an external relation. Important here is Putnam's doctrine of *convergence*. Our current scientific theories may be inadequate in various ways, but we can improve them or replace them with better theories, as we have in the past. What constitutes improvement is that theories come to approximate the truth better--truth in the non-intentional sense preferred by metaphysical realists. Theories converge on the truth, even if approximate truth is the best that can be hoped for in the case of an actual, non-ideal theory.

Part of what it means for theories to converge on the truth is for them to say more about the same things--*e.g.*, current theories of electricity tell us more about electricity than older ones did. Putnam explicates the point in "Explanation and Reference":

> [W]ith a few possible exceptions ... realists have held that there are successive scientific theories about the *same* things: about heat, about electricity, about electrons, and so forth; and this involves treating such terms as 'electricity' as *trans-theoretical* terms ... i.e. as terms that have the same reference in different theories.[7]

---

[4] Putnam (1975), p.270.

[5] See Chapter 2.

[6] I think that we can see "The Meaning of 'Meaning'" as an attempt to grant a great deal to the view that meaning is use without actually accepting that view and without abandoning metaphysical realism. In retrospect Putnam's conversion to internal realism is, if not predictable, not entirely surprising. A similar dialectic is at work below in my discussion of Burge, but Burge has not yet renounced --as I believe he should--his metaphysical realist ways.

[7] Putnam (1975), p.197.

As we saw in Chapter 3, the idea that homophonic terms of different theories co-refer is made dubious by the seeming conceptual relativism of thinkers like Kuhn. Putnam's worry is that the loss of coreferentiality undermines realism, for if the reference of theoretical terms shifts from one theory to the next, then we seem pushed toward an instrumentalist account of science. Each theory must regard its predecessor as lacking in some way, and if the meaning of theoretical terms shifts, then each theory must suppose that the entities admitted by its predecessor do not exist.

> [E]ventually the following meta-induction becomes overwhelmingly compelling: *just as no term used in the science of more than 50 ... years ago referred, so it will turn out that no term used now ... refers.*[8]

If no theoretical terms refer, then the idea that science tells us how the world is in itself is threatened. So, if one is sympathetic to (a) metaphysical realism and (b) scientific realism, then, claims Putnam, one should also be sympathetic to the idea of "trans-theoretical" terms. And if one has independent arguments for metaphysical and scientific realism, then one can hold that there *are* trans-theoretical terms, since this is implied by the combination of (a) and (b).

Now, treating reference as a component of meaning has consequences for semantic theories that view meaning as comprised of a speaker's intentions in uttering an expression, where such intentions are taken to be, *e.g.*, psychological states of the speaker. On this view, two speakers who are in the same kind of psychological state or, maybe, the same kind of brain state, utter terms with the same meanings, regardless of whether those terms have same reference. The speaker's psychological state constitutes meaning and determines reference, but reference is no part of meaning. Letting extension be an independent element of meaning, not completely determined by speaker's intentions requires the repudiation of the view that meanings are "in the head."

Putnam illustrates this point with a widely discussed and imitated example. Consider the inhabitants of Twin-Earth, a planet in a remote part of the universe, which is similar in many respects to Earth. They live in a world that is a phenomenological duplicate of our own, and have evolved, biologically and culturally, in ways parallel to our own. Indeed, they even use languages with the same phonemes and graphemes as our languages have, so that, were I conveyed to Twin-Earth, I could not immediately tell the difference.

But there are differences. For example, the substance that on Earth is called "water" by speakers of English has the chemical structure H_2O, but the substance that plays much the same role in the lives Twin-terrans and is called by speakers of Twenglish "water" has the (to us) exotic chemical structure XYZ. XYZ and H_2O share many properties and are equivalent for quotidian uses, but they behave differently in sophisticated laboratory tests. Now, suppose that I have a duplicate who resembles me in fine detail and resides on Twin-Earth. When he utters statements about water, he refers not to H_2O, but to XYZ, despite being, ostensibly, in the same psychological

---

[8] Putnam (1978), p.183f. See PMN Chapter VI for a perceptive discussion of this passage and related issues.

state[9] as I am, when I make statements about water and refer to H_2O. The substances named "water" differ from Earth to Twin-Earth; the reference of 'water' is not the same in English as in Twenglish. Since extension is part of meaning, 'water' does not mean the same thing on Earth as on Twin-Earth. Meanings "ain't in the head."[10]

Tyler Burge has used examples of this sort to argue that, like meaning, the individuation of "mental states"[11] must be accounted for "anti-individualistically." When Burge speaks of "individuation" he has in mind, not tokens, but types of intentional phenomena. So "individuation" has the sense of "classification." To individuate intentional phenomena is to say of what type a particular token is, and the types in question are narrower than beliefs, wishes, desires or intentions. What is at issue is not just whether an intentional token is a belief or a desire, but whether a particular belief is a belief about water, dogs, fire hydrants or Brian Mulroney's leg.

What is it to be an anti-individualist about the classification of intentional phenomena? A good way of bringing this out is to consider what it is to be an *individualist* about such classification. The psychological individualist thinks that, in Burge's vivid formulation, "the mind is somehow self-contained" (CEOS 118)--the nature of my beliefs and other intentional attitudes is quite independent of "the world," though many of those beliefs still pertain to the world. As Burge notes, no simple account of individualism does justice to the variety of philosophers who have held views that are individualistic in spirit, from Cartesians and idealists to present day reductive materialists, but a rough generalization will suffice: "[T]hey all think that the nature and individuation of an individual's mental

---

[9] Setting up the example is complicated by a number of factors that I shall ignore. We could not, for example, be in quite the same brain state--even if it were uncontroversial to identify psychological states with brain states--since his physiology would be based on XYZ and mine on H_2O. (Thanks to Tilman Lichter for this point.) And if we think of psychological states as functional states (Putnam's view at the time, but hardly uncontroversial), we still face the question of whether such states can be differentiated by a "narrow psychology" (*i.e.*, without reference to the speaker's environment) as Putnam assumes. This is a consequence of the view that Burge presents and which I will consider below, but a charitable way of reading Putnam would be as assuming this only for the sake of argument.

[10] Putnam (1975), p.227.

[11] I have reservations about the idea that beliefs and other "propositional attitudes" (a phrase I dislike) should be classified as "mental states"--let alone that such "mental states" should then also be identified, token by token, with tokens of physical states, or (worse) that the interpretation of a given "state" is given by its causal, functional relations with other states, or (still worse) that types of "mental states" should be identified with types of physical states, even intra-specially. But I shall not endeavour to take serious issue with these common assumptions here, though I shall try to avoid such talk where possible. In particular, I shall sometimes speak warily of "intentional phenomena" or "intentional attitudes."

kinds are 'in principle' independent of the nature and individuation of all aspects of the individual's environment" (CEOP 117).[12]

This view need not deny that my beliefs were, *e.g.*, contingently caused by features of my environment; it need merely assert that they did not have to be so produced, and that being produced in another fashion would not alter the kinds of beliefs that they are. A belief about water would be the same kind of belief, were it caused by H_2O or by some exotic chemical compound with many of the same properties as H_2O and indistinguishable from it for ordinary purposes.

Anti-individualism, as one might guess, denies individualism. The anti-individualistic thesis (AI) may be put as follows:

> AI The nature and individuation of an individual's intentional phenomena are dependent upon some aspect(s) of that individual's environment.

Burge offers a rough specification of *what* aspects of the environment are determinants of the nature and proper classification of a person's intentional phenomena. On his account what I believe depends partly upon what is in my environment about which I can have a belief.[13]

The application of Putnam's theory of meaning to Burge's account of the classification of intentional phenomena is clear. According to Putnam's example, the meaning of 'water' on Earth is not the same as the meaning of 'water' on Twin-Earth. Now, my beliefs are expressed in my linguistic behaviour; so, the interpretation of my beliefs depends upon the meanings of my words. It thus seems reasonable to conclude that what belief I have depends upon whether I inhabit Earth or Twin-Earth. When I have a belief about water, I have a belief about H_2O. When a Twin-terran has a belief about water, she has a belief about XYZ. Thus, we have different beliefs--or more precisely, different *kinds* of beliefs--even though we might express our beliefs in much the same way, and even though we might be in the same "psychological state" from a "narrow," or individualistic, point of view. The type of my belief, therefore, is at least partly determined by the environment that I inhabit. This conclusion rules out psychological individualism, for it supposes that a difference in the production of beliefs can make them different kinds of belief, despite the fact that they would be similarly expressed and dependent on phenomenologically similar perceptions. Meanings--even of beliefs--"just ain't in the head." The first assumption in my argument against Burge then is:

> (1) Burge holds AI.

---

[12] Burge offers a more detailed sketch, but the rough version seems clear enough for present purposes.

[13] It also depends on my membership in a linguistic community. See Burge (1979) for an extended treatment.

II. Anti-Individualism and First-Person Authority

What has all this to do with self-knowledge? The answer is that an anti-individualistic account of the individuation of belief-types seems to undermine the common presumption that an individual "knows her own mind" better than others do. In particular, Burge's anti-individualism seems inconsistent with what Burge calls the "Cartesian conception" (ISK 649) (or C, for short) of first-person authority. The thesis of first-person authority is roughly the claim that a person knows her own mind better than others do (and better than she knows the minds of others). So, one can embrace a version of first-person authority without accepting C. Davidson supposes, for instance, that a person cannot be massively and systematically mistaken about her own beliefs, desires, etc., but also holds that on particular occasions a person can be mistaken about her own intentional phenomena.[14] As a minimal requirement for first-person authority, I propose M1: I am not generally mistaken about my actual intentional phenomena. That this "weak authority"--as I shall say--is a minimal constraint on first-person authority is the second assumption of my critique of Burge:

(2) First-person authority requires at least M1.

Conversely, one might hold that authority is stronger--that a person is never mistaken about her beliefs, desires, hopes, etc. or, indeed, that she simply *cannot* be so mistaken (perhaps with the proviso that she must be sufficiently attentive in order to avoid error, but that such attention is possible). This is a first approximation of the "strong authority" that I shall take to characterize C--though, as we saw in Chapter 6, it may oversimplify Descartes' own views.[15]

Now, if I cannot be mistaken about what I believe, then Burge's version of AI would seem false for the following reason: if my beliefs are partly individuated by my environment, then Twin-Earth examples suggest that my environment might remain phenomenologically similar, while the nature of my beliefs changed. We might tell a science-fiction story in which I am kidnapped in my sleep by the unfriendly crew of a passing Vogon spaceship and switched with my duplicate on Twin-Earth. I wake up, ignorant of my new situation, and remain there for many years.[16] During that time, according to Burge, I acquire beliefs about XYZ, but do not regard myself to have undergone a change of belief about water. If Burge is right, I am mistaken about some of my own beliefs, since their nature depends on my environment. And the greater the number of deep differences that underlie the superficial similarities between Earth and its Twin, the more wildly and widely I can be mistaken about my own intentional phenomena. "Our problem," then, as Burge says, is

> ... understanding how we can know some of our mental events in a direct, nonempirical manner, when those events depend

---

[14] See FPA, Davidson (1987) and Davidson (1989).

[15] Burge acknowledges this interpretive point. See ISK 651n4.

[16] *Cf.* ISK 652.

> for their identities on our relations to the environment. (ISK 650)

If Burge is to reconcile his anti-individualism with first-person authority, then he must find a weaker version than C.

To understand his proposal for doing so, we must re-examine C. The claim that I am strongly authoritative about my intentional phenomena can be broken down into three clauses:

C1 I am *actually* authoritative about my actual intentional phenomena.
C2 In any counterfactual situation I *would be* authoritative about my intentional phenomena in that situation.
C3 I am *actually* authoritative about my intentional phenomena in any counterfactual situation.[17]

But how should we understand "authority" here? Since this is supposed to capture the *Cartesian* conception of first-person authority, it ought to be construed as "strong authority." Moreover, the point of distinguishing C1 from C2--if there be any--is lost if we take strong authority simply as the claim that I *cannot* be mistaken. Since Burge draws the distinction (see CEOP 123f), it will be simpler to follow him by treating strong authority as amounting to my *never* being mistaken about that regarding which I am authoritative. This gives us:

C1.1 I am actually never mistaken about my actual intentional phenomena.
C2.1 In any counterfactual situation I would never be mistaken about my intentional phenomena in that situation.
C3.1 I am actually never mistaken about my intentional phenomena in any counterfactual situation.[18]

---

[17] We might add a further clause: C4 In any possible situation I would be authoritative about my intentional phenomena in any possible situation. C4 entails C3, but the reverse entailment does not hold.

[18] If we treat "authority" as the claim that I *cannot* be mistaken about that regarding which I am authoritative, then we have:

C1.2 I actually *cannot* be mistaken about my actual intentional phenomena.
C2.2 In any counterfactual situation I *could not* be mistaken about my intentional phenomena in that situation.
C3.2 I actually *cannot* be mistaken about my intentional phenomena in any counterfactual situation.

Depending upon how we construe "accessibility" relations amongst "possible worlds," C1.2 could be taken to imply C2.2. That is--to say that I cannot be mistaken about *p* is to say that in any world accessible to the actual world I *would not* be mistaken about *p*, and if any world that is accessible to worlds that are in turn accessible to the actual world is also itself accessible to the actual world, then C1.2 --> C2.2. Under such conditions C1.2-C3.2 reduce to C1.1-C3.1. Since Burge distinguishes C1 from C2 (see CEOP 123f), it makes better sense of his views to treat "authoritative" as "never mistaken."

What is wrong with C and its variants? I suggested in Chapter 6 that something like C arises from dichotomizing the objective and the subjective. Although an emphasis on certainty helps explain the Cartesian version of first-person authority, such certainty does not seem possible (or necessary), unless one treats sceptical hypotheses as real possibilities--that is, unless one adopts metaphysical realism, explicitly or implicitly (as the relativist does).

Burge's suggestion is similar but not quite the same. He holds that Cartesian scepticism contributes to C by encouraging us to think that we know what our intentional phenomena would be in counterfactual situations. This is especially so, given the emphasis that external-world scepticism places on individual subjectivity:

> From the 'inside', from a 'first-person' point of view, one develops an impression of the independence of the nature of one's mental life from outside determining factors. One has a vivid sense of how the world seems; but one remains conscious of the contingency of the relation between the way the objective world is and its effects on us. That is, the same sensory effects could seemingly have been systematically produced by a variety of different sets of cause-cum-laws. Then our vividly grasped thoughts would be mistaken. These sorts of consideration have led many to conclude that individualism must be true. (CEOP 120f)

But the Cartesian gets herself into trouble right here on Burge's subtle account. According to him, the Cartesian ends up with C, because she makes a mistake about the evaluation of counterfactual conditionals. She conflates the evaluation-conditions for two different kinds of counterfactual conditionals: those pertaining to the possibility of Cartesian scepticism and those pertaining to the individuation of intentional phenomena. Burge writes,

> It is a well-known point that in considering counterfactual situations we hold constant the interpretation of the language whose sentences we are evaluating in the counterfactual situations. It is quite possible to consider the truth or falsity of interpreted sentences even in counterfactual situations where those sentences could not be used or understood. Similarly for our thoughts when we are considering the Cartesian situations. We hold our thoughts constant. We consider situations in which the thoughts that we have would be false. And we concede that we could in principle be mistaken in thinking that the world is not arranged in one of the ways that would make our thoughts radically false. (CEOP 122f)

But, says Burge, to answer a question like: What thoughts would we have if our epistemic relation to the world were radically different from what we take it to be?--we cannot hold the interpretation of our thoughts constant. That interpretation is just what we want to settle; so, it would be question-begging to decide it beforehand. On Burge's

---

[18] Also, bear in mind that slightly more sophisticated versions can be obtained by specifying further that I must have considered my intentional phenomena with sufficient attentiveness.

view, the Cartesian cannot insist from the outset that the "possible world" at which we evaluate the sceptic's hypothesis is the same world at which to interpret our beliefs. Is it compatible with our having all our actual beliefs that we should be brains in a vat? Yes, thinks Burge. But if we *were* brains in a vat, then we would have beliefs, not about brains and vats, but about illusory brains and vats, or computer-generated electrical impulses.

The conflation of these issues, according to Burge, is a source of the tension between AI and first-person authority. If my belief-types are individuated by circumstances "external" to me, circumstances about which *I could be mistaken*, and if I could be mistaken in all my beliefs about the empirical world, then does it not follow that I could be mistaken in all of my beliefs about my belief-types? No, responds Burge, because our lack of authority about what beliefs we *would* have in the sceptic's counterfactual scenarios, does not entail any lack of authority about the beliefs that we in fact *do* have.

> First-person authority presupposes our thoughts as given; we are then authoritative about those thoughts. But our thoughts are determined to be what they are partly by the nature of the environment. And we are authoritative about neither our environment nor the nature of that determination. (CEOP 124)

The Cartesian individualist, thinks Burge, insists (1) that she is authoritative about the types of her actual empirical beliefs, (2) that she would be authoritative about the types of her counterfactual beliefs, *and* (3) that she is *actually* authoritative about the types of her counterfactual beliefs. This is just a reiteration of C. And to reconcile first-person authority with anti-individualism, says Burge, we should reject C3. But he retains the first two Cartesian theses:

> We are authoritative about some of our actual thoughts about the empirical world; and we can imagine those very thoughts being quite mistaken. Moreover, whatever our thoughts would be if the counterfactual situation were to obtain, we would be authoritative about them. But we are *not* authoritative about what our thoughts about the empirical world would be if the counterfactual cases were actual. (CEOP 123f)

This gives us Burge's conception (B) of first-person authority:

B1 (C1) I am *actually* authoritative about my actual intentional phenomena.
B2 (C2) In any counterfactual situation I *would be* authoritative about my intentional phenomena in that situation.

If we again take "authoritative" to mean "never mistaken," then we also have an explication of B:

B1.1 (C1.1) I am actually never mistaken about my actual intentional phenomena.
B2.1 (C2.1) In any counterfactual situation I would never be mistaken about my intentional phenomena in that situation.

This provides my third major assumption:

(3) Burge holds C1.1.

By so weakening C, claims Burge, it becomes possible to reconcile AI with first-person authority in a way that the Cartesian had ruled out.

But how am I authoritative about my beliefs, if I could be utterly mistaken about what individuates them into types? Burge offers this reply. "Basic self-knowledge" is the title he gives to the kind of self-knowledge that is embodied in judgments like "I am thinking about Vienna," or "I think that water is a liquid":

> In basic self-knowledge, a person does individuate his thoughts in the sense that he knows the thought tokens as the thought tokens, and types, that they are. We know which thoughts we think. When I currently and consciously think that water is a liquid, I typically know that I think that water is a liquid. So much is clear. (ISK 653)

But how does this individuation by the thinker differ from the individuation of beliefs by her environment? How is it that she need not know the empirical conditions that individuate her beliefs in order to "individuate" them in some other sense of the word?

Burge suggests that we distinguish between conditions that "enable" a belief, or make it possible, and conditions that justify it. Take, first, an ordinary empirical belief, rather than a first-person belief. I cannot have beliefs about water, unless I have had some sort of interaction with water, or with other persons who have had such interaction or have been told about such interaction and can describe it to me. This set of conditions is here very simply described, but there are--it seems--many particular things that would have to be the case in order for me to have beliefs about water. The description can become much more complicated. Still, I need not know about all those conditions in order to have beliefs of that type:

> To know that water exists, or that what one is touching is water, one cannot circumvent empirical procedures. But to *think* that water is a liquid, one need not *know* the complex conditions that must obtain if one is to think that thought. Such conditions need only be presupposed. (ISK 654)

Much the same holds true, Burge thinks, for first-person beliefs--*i.e.*, beliefs about one's own intentional phenomena. For me to believe that I am thinking about water many conditions must obtain. Among them will be ones that are necessary to my thinking about water. And this is no mere coincidence--it is not even a well-established empirical regularity. Rather, it is a logical connection, according to Burge:

> When one knows that one is thinking that *p*, one is not taking one's thought (or thinking) that *p* merely as an object. One is thinking that *p* in the very event of thinking knowledge-ably that one is thinking it. It is thought and thought about in the same mental act. So any conditions that are necessary to thinking that *p* will be equally necessary to the relevant knowledge that one is thinking that *p*. Here, again, to think the thought, one need not know the enabling conditions. It is enough that they actually be satisfied. (ISK 654)

To think I am thinking of water is to be thinking of water; so, to think I am thinking of water is to know that I am thinking of water. "Crudely put," says Burge, "our knowledge of our own thoughts is immediate, not discursive" (ISK 656). This immediacy disappears when we consider our actual knowledge of counterfactual beliefs.

III. Real Possibility: A Problem for Burge

I believe that Burge's attempted reconciliation of AI and first-person authority will not work, because I think that his critique of C fails, and that his view differs insufficiently from C to be coherent. I shall devote most of my attention to the first criticism.

At first glance, it might seem that I should be sympathetic to Burge's approach. Some of his remarks are reminiscent of my complaint in Chapter 2 that the metaphysical realist wants to situate herself simultaneously both inside and outside the sceptic's vat. Cartesian individualism seems plausible, he says, because we tend to confuse first- and third-person perspectives. We begin with a first-person judgment: "we think that we are thinking that water is liquid." But envisioning the Cartesian's sceptical scenario requires us to switch to the third-person, "a perspective on ourselves from the outside." Once we have done this, Burge continues,

> ... we are easily but illegitimately seduced into the worry that our original first-person judgment is poorly justified unless it can somehow encompass the third-person perspective, or unless the third-person perspective on empirical matters is irrelevant to the character of the first-person judgment. In this fallen state, we are left with little else but a distorted conception of self-knowledge and a return to individualism. (ISK 661)

This vivid description sounds like the sort of analysis that I have given of Cartesian scepticism in terms of "real possibility." The metaphysical realist, I suggested, is tempted to object to Putnam's argument that we could not really always have been brains in a vat by saying that although we could not knowingly say that we have always been brains in a vat, we could yet be in a situation that we cannot express. Such a criticism, I argued, treats reference as a non-normative notion and, indeed, even flouts any plausible version of so-called causal theories of reference. The metaphysical realist mistakes the logical possibility of my always having been a brain in a vat for the alleged real possibility of the same. This seems rather like confusing the third-person perspective with the first-person.

But there is a good reason for thinking that Burge has a different point in mind, despite its *prima facie* similarity. The reason is that he is a metaphysical realist himself, and I shall now argue that my account of Cartesian scepticism shows Burge's critique to be flawed.

The story that I told about Cartesian scepticism in Chapter 2 went something like this. The sceptic thinks that in some sense of the word it is *possible* that I have always been a brain in a vat. It seems that such deception is logically possible; to say that I have always been a brain in a vat involves no contradiction. But the logical possibility that *p* is not generally a reason for thinking that *p* or even that probably *p*. So, for the sceptic's "doubts" to be compelling, they must rest on something stronger than just logical possibility.

Burge acknowledges this, when he writes that the doubts of the Cartesian sceptic constitute not just a logical, but an *epistemic* possibility (CEOP 122). But we saw that this is not right either--at least, not if an epistemic possibility is one that is compatible with everything that I know. To say that, given everything I know, I might always have been a brain in a vat is to assume that I do not know that I have not always been a brain in a vat. If the sceptic's hypothesis is intended as an epistemic possibility, then it begs the question.

But there is another sense of "possibility" that does better justice to the intentions of the sceptic. The sceptic claims that it is a real possibility that we are completely mistaken about the way the world is. A real possibility, recall, is a possibility whose actuality would explain some actual phenomenon or phenomena of which we are assumed to have knowledge. What we are assumed to know here is what seems to be the case about the world. That I have always been deceived by an evil demon, or that I have always been a brain in a vat, the sceptic says, is just as good an explanation for my intentional phenomena. All the marks of evidence that we have regarding the world around us would be explained just as well by the hypothesis that we have always been brains in a vat. Systematic delusion is not just logically possible, says the sceptic, but really possible. Moreover, she continues, if two rival hypotheses give equally good explanations for some class of phenomena, then I am no more justified believing one than believing the other. If knowledge of the "external" world requires that the external-world hypothesis have better justification, then I cannot correctly claim to have knowledge of a world beyond my senses.

As we saw, four likely strategies present themselves as ways of responding to the sceptic, so understood.

> (i) Argue that the sceptic's account of justification or knowledge is faulty, so that I could be justified in believing in the "external" world, despite the explanatory power of the sceptical hypothesis.
>
> (ii) Deny that knowledge requires justification at all.
>
> (iii) Argue that although the Cartesian-sceptical situation (CSS) is a real possibility, it is not as good an explanation of intentional phenomena as the external-world hypothesis.
>
> (iv) Deny that the CSS is a real possibility.

The last of these, remember, is out of the metaphysical realist's reach, because metaphysical realism implies that the nature and existence of the world are independent of our abilities to know or describe them, and the only plausible construal of this independence is as saying that the sceptical scenario is really possible.

Now, Burge *explicitly rejects* strategy (iv) and, so, internal or pragmatic realism. In "Cartesian Error and the Objectivity of Perception," he gives qualified support to "the oft-repeated slogan that error presupposes a background of veridicality," but continues,

> I think that this slogan is sometimes misused. I think that we are not immune from fairly dramatic and wholesale error in characterizing the nature of the empirical world. (CEOP 130f)

"Quine and Davidson," he notes, "sometimes use this important idea with insufficient discrimination" (CEOP 131n4). His own "qualified" support for this slogan stems from his conviction that intentional phenomena are to be individuated non-individualistically. If I seem to see small shadows on the side of a brick-building, then I typically do see them, because I would not have the perceptual type "seeming to see small shadows," if there were no, or very few, small shadows in my environment. This thesis, however, he finds compatible with "fairly dramatic and wholesale error" for reasons considered in the preceding section--the Cartesian confuses the evaluation-conditions for two sorts of counterfactual-conditional claims, and I need not know the enabling conditions of my intentional phenomena.

"Individualism and Self-Knowledge" also rejects strategy (iv) thus:

> Several philosophers have thought that anti-individualism, combined with the view that we are authoritative about what thoughts we think, provides a "transcendental" response to scepticism. ... I believe, however, that there is no easy argument against scepticism from anti-individualism and authoritative self-knowledge. (ISK 655n6)

Here the focus of criticism is Putnam, whose arguments "do not do much to undermine scepticism," he says, endorsing Brueckner's objections.[19] Burge supports a version of strategy (i)--arguing that the sceptic's account of justification is faulty. The critic of scepticism, he holds, should "deny[] that perceptual knowledge must be justified by separately insuring that the enabling conditions [for perceptual knowledge] hold and the sceptic's defeating conditions do not hold" (ISK 655f). That is, the sceptic puts unreasonable demands on the concept of justified belief and, hence, on the concept of knowledge. Just as I need not know the enabling conditions of my intentional phenomena to know what my intentional phenomena are, according to Burge, neither need I know the enabling conditions of my perceptual beliefs to be justified in holding them. "It is a fundamental mistake," he says, "to think that perceptual knowledge of physical entities requires, as a precondition, knowledge of the conditions that make such knowledge possible" (ISK 654).

Despite some tempting parallels between Burge's position and my own, then, there are important orthogonalities. And these differences of philosophical direction pose a worry for Burge: real possibility may be distinct from epistemic possibility as described in Chapter 2, but it does have an epistemic component. A real possibility is one whose actuality would explain some actual phenomenon of which we are *assumed* to have knowledge. What we are assumed to have knowledge of here is what seems to be the case about the world. But *given* that we can individuate our beliefs about the world, and given (were the sceptic right) that we might really be mistaken in holding those beliefs, it follows that we might be completely mistaken about the external world and still know our own beliefs, if we know them at all. *Therefore, for at least a range of counterfactual situations--real possibilities that*

---

[19] Brueckner's view is a version of the criticism I recounted above and rebutted in Chapter 2: *i.e.*, Putnam has shown only that we cannot knowingly say that we have always been brains in a vat, but we could be for all that. See Brueckner (1986).

are not actual--we know what our beliefs would be in those situations, if scepticism is a real possibility. If our sceptic holds that we have authority about what beliefs we have concerning the world, then we have authority about our really possible beliefs. If scepticism is a real possibility then, if C1.1 is true, C3.1 is also true. And there is no reason to take the Cartesian to claim similar authority about counterfactual beliefs in merely logically (or physically) possible situations. The individualism embodied in the *real* necessity of my knowing my own beliefs is enough. This is to say that the Cartesian picture of first-person authority should yet be amended as follows:

C1.3 (C1.1) I am actually never mistaken about my actual intentional phenomena.
C2.3 In any *really possible* counterfactual situation I would never be mistaken about my intentional phenomena in that situation.
C3.3 I am actually never mistaken about my intentional phenomena in any *really possible* counterfactual situation.

Three major points need to be made. First, C1.3 is just C1.1 again, and as before, if C1.3 is true, C3.3 is also true--if I am never mistaken about my actual intentional phenomena, then I am actually never mistaken about my intentional phenomena in really possible counterfactual situations, given that the CSS is really possible. So:

(4) If the CSS is really possible, then C1.1 --> C3.3.

Secondly, we have seen that Burge denies the authority of my actual beliefs about my counterfactual beliefs. But if Burge is merely saying that I am not authoritative about what beliefs I could *logically* have, his view does not differ much from the Cartesian's, represented by C1.1-C3.3. So we must, I think, take him to reject the claim that I actually know what my beliefs would be in every really possible counterfactual situation (C3.3). His reasons for doing so provide my fifth assumption--anti-individualism is incompatible with my actual authority about my really possible intentional phenomena:

(5) AI --> ~C3.3.

Thirdly, while the range of possibility in C1.3-C3.3 is narrower than logical possibility, it remains a very strong version of first-person authority. This is clear if we examine a non-Cartesian version:

M1 I am not generally mistaken about my actual intentional phenomena.
M2 In any really possible counterfactual situation I would not be generally mistaken about my intentional phenomena in that situation.
M3 I am not generally mistaken about my intentional phenomena in any really possible counterfactual situation.

M1, I suggested in the preceding section, is a minimal constraint on first-person authority. M2 and M3 I take to be plausible extensions of this minimal account. While M1-M3 give a non-Cartesian account of first-person authority, we can still imagine someone taking this view and holding the CSS to be really possible. One can have worries about Cartesian scepticism without thinking that basic self-knowledge claims

are incorrigible. However, this extended minimal account shares two features with stronger versions: first, if the CSS is really possible, then if I am not generally mistaken about my actual intentional phenomena, neither am I actually, generally mistaken about my really possible intentional phenomena--my intentional phenomena in really possible counterfactual situations, including the CSS. This gives us:

(6) If the CSS is really possible, then M1 --> M3.

Secondly, and as a consequence of this first point, if the CSS is a really possible, then M3 implies the individualism that Burge rejects, for if the CSS is a real possibility, then M3 implies that I have weak authority about my really possible intentional phenomena. From this I get my final major assumption: if the CSS is really possible, then anti-individualism is incompatible with my actually having even weak authority about my really possible intentional phenomena:

(7) If the CSS is really possible, then AI --> ~M3.

Now, what are the consequences of these observations for Burge's position? I listed above four kinds of anti-sceptical responses. Burge's stated views are congenial to strategy (i), as we have seen. But if his argument against C is to stick, I submit that he must use strategy (iv)--*i.e.*, he must abandon metaphysical realism.

Why is strategy (iv) the only one to which Burge can appeal to reconcile first-person authority with AI? The problem is that if the CSS is really possible, then, first, anti-individualism is not compatible with my being either strongly or weakly authoritative about my intentional phenomena in really possible counterfactual situations, and secondly, my having such authority is implied by my authority about my actual intentional phenomena. If Burge rejects C3.3 and M3, then--assuming that he wants to remain a metaphysical realist--he must also reject C1.1 and M1.[20] That is, he must reject both strong and weak versions of the claim that I have authority about my actual intentional phenomena. Rejecting C1.1 would require a change in Burge's views, since he thinks I am strongly authoritative about my actual intentional attitudes. But--more seriously--rejecting M1 amounts to rejecting a minimal constraint on any account of first-person authority. As long as Burge holds that the CSS is really possible, his attempt must fail. Metaphysical realism, anti-individualism and first-person authority cannot all be held at once.

Let me summarize what I think is a difficult argument:

(1) Burge is an anti-individualist--he holds that the classification of an individual's intentional phenomena depends upon aspects of her environment. (2) If Burge is to admit first-person authority, he must accept at least (M1) that I am not generally mistaken about the classification of my actual intentional phenomena. (3) In fact, Burge holds a stronger claim (C1.1) that I am never mistaken about my actual intentional phenomena. But he denies (C3.3) that I am actually never wrong about my intentional phenomena in really possible counterfactual

---

[20] If metaphysical realism is rejected, then the counterfactual situations raised by the sceptic do not satisfy the requirements of M3. Thus, that M1 entails M3 can be accepted by someone who rejects the real possibility of the situation hypothesized by the sceptic.

situations, because he holds (5) that anti-individualism is not compatible with such epistemic authority. I have argued (4) and (6) that if the Cartesian sceptical scenario is really possible, then my strong or weak authority about my actual intentional phenomena implies my (respectively) strong or weak authority about my really possible intentional phenomena. I have also argued (7) that if metaphysical realism is the case, then anti-individualism is incompatible with the claim (M3) that I am generally not mistaken about my intentional phenomena in really possible counterfactual situations. If, with Burge, we assume (8) that metaphysical realism is correct, then (9) my authority (weak or strong) about my actual intentional phenomena extends with similar force to my knowledge of my intentional phenomena in really possible counterfactual situations. Therefore, by *modus tollens*, (10) if I am not authoritative, either weakly or strongly, about my intentional phenomena in really possible counterfactual situations, then I am not authoritative--weakly or strongly--about my actual intentional phenomena. Because Burge is an anti-individualist, (11) he should hold that I'm not authoritative, weakly or strongly, about my intentional phenomena in really possible counterfactual situations. Therefore, (12) he should also hold that I have no authority, weak or strong, about my actual intentional phenomena. But Burge thinks that I have strong authority about my actual intentional phenomena. Therefore, (14) his position is inconsistent, and (15) if he remains a metaphysical realist and an anti-individualist, then he must reject first-person authority.[21]

Burge's elegant response to the Cartesian individualist, I conclude, involves a misdiagnosis. The problem is not that the Cartesian sceptic thinks that she is actually authoritative about the beliefs she would have in every counterfactual situation. Rather, it is that she wants to evaluate a *really possible* counterfactual as she would a *logically* or *physically* possible counterfactual.

---

[21] The proof-structure can be quasi-formally represented thus:

1. Burge holds AI. [Burge]
2. First-person authority requires at least M1. [Hymers]
3. Burge holds C1.1. [Burge]
4. If the CSS is really possible, then C1.1 --> C3.3. [Hymers]
5. AI --> ~C3.3. [Burge]
6. If the CSS is really possible, then M1 --> M3. [Hymers]
7. If the CSS is really possible, then AI --> ~M3. [Hymers]
8. ¦Assume that the CSS is really possible.
9. ¦Then C1.1 --> C3.3 and M1 --> M3. [4,6,8]
10. ¦So, ~C3.3 --> ~C1.1 and ~M3 --> ~M1.
11. ¦Burge should hold ~C3.3 and ~M3. [1,5,7,8]
12. ¦Therefore, Burge should reject C1.1 and M1. [10,11]
13. ¦But Burge endorses C1.1. [3]
14. Therefore, Burge's position is inconsistent, if he holds that the CSS is really possible. [8-13]
15. Moreover, Burge must reject first-person authority altogether, if he holds that the CSS is really possible. [2,8-12]

If Burge wants to hold AI and first-person authority, then he cannot accept 8--he must reject metaphysical realism.

So much, then, for my major criticism of Burge's views: his critique of C fails, because his account of Cartesian scepticism neglects the importance of real possibility. But I shall now show that his position (B) differs insufficiently from C to be coherent. This should not be surprising, given the connections between C and Cartesian scepticism on the one hand and metaphysical realism and scepticism on the other.

Burge retains the view that, given my beliefs about the world, I have an authority about them that implies the incorrigibility of my judgments of basic self-knowledge. All he denies is that I also have actual authority about my beliefs in really possible counterfactual situations. He thus thinks that some of my judgments of basic self-knowledge are corrigible--*viz.*, those about beliefs I would have in a really possible counterfactual situation (though he might not put the point this way). But his attitude toward C1.1 and C2.1 does not differ from the Cartesian's--C1.1 and C2.1 are B1.1 and B2.1.

Now, if I am strongly authoritative about basic self-knowledge, then there is no difference between *being right* in my first-person beliefs and just *having* those first-person beliefs. Burge tells us as much: "One knows one's thought to be what it is simply by thinking it while exercising second-order, self-ascriptive powers" (ISK 656). But if thinking one is "right" does not differ from being "right," then why should I not have strong authority about what beliefs I would have in really possible situations? If thinking I am right does not differ from being right, then "here we can't talk about 'right'" (PI §258)--there is no fact about what beliefs I have. But if so, why should there actually be a fact about what beliefs I *would* have? There would not be such a fact counterfactually, since I would then have strong authority; so, how is it that that which would be indeterminate, if actual, is counterfactually determinate? Or is it that in both cases there is no fact, but in one I have some surd authority that is surdly absent in the other? If I, as a language-user, understand a concept, then I know what *would* count as an instance of it, not just what does. If I know only the latter, then I have only a partial grasp of the concept, and it is odd to say that I have authority. If I can be wrong about what would count, surely I can be wrong about what does.[22]

So, if only I can say which phenomena accord with type T, then only I can say what really *would* accord with type T, and I can interpret T as I wish. To assess counterfactuals, we fix "the interpretation of the language whose sentences we are evaluating in the counterfactual situations" (CEOP 122), says Burge. But only I can say what counts as holding the interpretation of my first-person beliefs constant. Only I can say whether my really possible belief would be different. *I* make the "rules" here, and I interpret them as I see fit.

Burge claims that anti-individualism can be reconciled with first-person authority, because we need not know what conditions enable a belief in order to have the belief. If basic self-knowledge is self-verifying, we need not know what enables that knowledge in order to have that knowledge. But even if we accept Burge's anti-sceptical ploy (i), there are some enabling factors that I must know in order to have basic self-knowledge. *I must, by and large, understand my own words.*

This requirement has stood behind my arguments against seeing metaphysical realism and relativism as the only epistemological games in town. And it poses a problem for Burge in the following way. "My

---

[22] Thanks to Robert Bright for pressing me on this point.

knowledge that I am thinking that mercury is an element," he says in a footnote, "depends on an ability to think--not explicate--the thought that mercury is an element" (ISK 661). But to be thinking that mercury is an element, I must be able to say *something* about what mercury is or what an element is. Maybe I need not have had training in modern chemical theory, but I do need a sense of the grammar of these words. I need to know that 'mercury' and 'element' are both standardly used as mass-terms, rather than count-terms. Or if this picture seems over-intellectualized, it might be safer to say that I need to know that 'mercury' is more like 'water' than like 'dog'--it names of a kind of "stuff." I do not understand my language one word at a time in total isolation from its other words. To the extent that I do not understand the uses of the words 'mercury' and 'element' (and 'is' and 'and') I am not "thinking a thought" at all, but making noises or calling sounds to mind as I might remember a fragment of music. The standard use of a word is a matter about which I can be wrong and about which I can have a confused or partial understanding. *A fortiori* I can have in specific cases a confused or partial understanding of what I believe. But this conflicts with Burge's version of first-person authority.

Philosophical discourse shows how I *can* be wrong about what I think, because I only partially understand the concepts I use. Making an argument, I can suddenly see that I mistook the logic of the concepts involved. What I thought followed from my premises no longer seems validly linked to them. I did not understand what it was I was thinking. This is not a case of *basic* self-knowledge or belief, but there is no clear line between basic and not-so-basic self-belief.

IV. An Alternative Picture

Is there is no difference between my knowledge of myself and my knowledge of others? I think that there are cases in which knowledge of others is very like typical cases of self-knowledge. Conversely, there are instances in which knowledge of oneself is very like typical cases of knowledge of others. Finally, there are many cases in which knowledge of oneself is very different from knowledge of others. But, further, I believe that such differences are of degree, not of kind. In what follows I shall join Donald Davidson in arguing that the special character of self-knowledge can be captured by drawing out some consequences of a claim to which I have often had recourse: that I cannot be completely ignorant about the meanings of my own words, but I can be mistaken in particular cases.[23] I shall also argue, however, that this special character is not absolute, because my knowledge of others sometimes resembles my self-knowledge.

In a number of papers Davidson has argued that there is an asymmetry between the authority of self-knowledge claims and the authority of knowledge-claims about others, and that this asymmetry is best grasped and explained by focusing, not on epistemological considerations about warrant, but on "relations between agents and utterances" (FPA 109).[24]

---

[23] To this I will add in Chapter 8 that there is a sense in which my ability to say what my thoughts and feelings are is important to my capacity for agency, because in articulating my beliefs and desires I may come to view them differently, as Charles Taylor has suggested.

[24] See also Davidson (1987) and Davidson (1989).

Davidson thinks it "obvious" that there is an asymmetry between "first person present tense claims about [propositional] attitudes, and other person or other tense claims" (FPA 104). But he also urges that any account of first-person authority avoid subjectivism by recognizing that I can be mistaken in my first-person judgments: "Error is possible; so is doubt" (FPA 103). The two points are closely related.

Is it "obvious" that self-knowledge differs from knowledge of other minds? One might hold the contrary, that there is no difference in the way I have knowledge of myself and the way I have knowledge of others. We might take the difference to be like the difference between one's ability to know what a close friend or lover is thinking and one's tendency to make mistakes about the thoughts of casual acquaintances and to miss the mark utterly with strangers. The reliability of one's understanding falls off as one's familiarity with the individual and her socio-cultural circumstances falls off. I know my own mind better than other minds, because I spend more time with myself, but I know by interpretation and inference, just as I know the minds of others.

Davidson attributes roughly this position to Gilbert Ryle. In *The Concept of Mind* Ryle, arguing against "Privileged Access"--a strong version of first-person authority--suggested that my self-knowledge is of precisely the same character as my knowledge of others:

> I learn that a certain pupil of mine is lazy, ambitious and witty by following his work, noticing his excuses, listening to his conversation and comparing his performances with those of others. Nor does it make any important difference if I happen myself to be that pupil. I can indeed then listen to more of his conversations, as I am the addressee of his unspoken soliloquies; I notice more of his excuses, as I am never absent, when they are made.[25]

Davidson thinks that Ryle makes a mistake about the nature of the alleged "asymmetry." The fact, he argues, that my self-knowledge is not based on a special power of introspection, or on my ability to be "thinking that *p* in the very event of thinking knowledgeably that one is thinking it," as Burge says, does not show an absence of asymmetry in the *authority* attributed to such judgments, in contrast with other kinds of knowledge-claims. The Rylean position Davidson says, citing the similar views of Ayer, amounts to likening self-knowledge to an eye-witness report and knowledge of others to second-hand news. As such, he claims, it falls short of explaining the asymmetry:

> [F]irst person attributions are not based on better evidence, but often, at least, on no evidence at all. The authority of the eyewitness is at best based on inductive probabilities easily overridden in particular cases: an eyewitness is discredited and his evidence discounted if he is a notoriously unreliable observer, prejudiced, or myopic. But a person never loses his special claim to be right about his own attitudes, even when his claim is challenged or overturned. (FPA 104)

---

[25] Ryle (1949), p.169.

I usually do not rely on evidence to decide what I believe about, *e.g.*, everyday objects that surround me. I don't have to check to see whether I believe there is milk in the fridge or coffee in the canister, any more than I usually have to examine the position of my limbs to determine whether I am walking or sitting, or writing or making pastry. Annette Baier puts the point succinctly, when she says, "If one has the know-how, one knows one has it and so knows that, in normal conditions, one's act succeeds. Practical knowledge is the human approximation of divine knowledge"[26]--but only an approximation. That knowledge of my own beliefs about, *e.g.*, household objects is like practical knowledge, is not explicitly suggested by Davidson, but inasmuch as linguistic understanding is, as Baker and Hacker say, "*akin* to an ability,"[27] the comparison is not simply gratuitous.

But what of Davidson's further point that a person "never loses" her first-person authority? I think that there are cases in which this can happen, cases that might sometimes be regarded as pathologies, but which are no less real for that. I am thinking of such phenomena as self-deception, personality disorders, and ideological delusion. For Davidson, such phenomena--at least, self-deception and the Freudian unconscious--may be real enough, but he says, they do not "threaten[ ] the importance of first person authority" (FPA 105). On this much we can agree, but it does not justify the claim that a person *never* loses her first-person authority. There are clearly cases in which self-knowledge is likely to be interpretive and inferential.[28]

I suggested above that we compare knowledge of one's beliefs with practical knowledge. The comparison captures much of Davidson's point: my knowledge of my own beliefs and desires, hopes and expectations is intimately linked with my linguistic abilities. Davidson's complaint with some of the positions he examines is that those who recognize that there is an asymmetry between first- and other-person ascriptions of beliefs and the like tend merely to "restate the asymmetry" (FPA 101), without offering an explanation for it.[29] However, he contends, if we examine a person's capacity to learn and use languages, an explanation of the asymmetry is forthcoming--one that turns on the fact that a speaker cannot be utterly mistaken about the meanings of her own words. What distinguishes my knowledge of my beliefs and desires from my knowledge of another's is that, since I typically know what my words mean, I typically know what beliefs, desires, hopes, etc. they express when I utter them. Says Davidson, "the assumption that I know what I mean necessarily gives me, but not you, knowledge of what belief I expressed by my utterance" (FPA 110).

We need to make two things clear: the reason that my linguistic knowledge gives me knowledge of my beliefs etc., and the reason that it supposedly does not give another such knowledge with the same reliability. The first point is easy enough to understand. For me to

---

[26] Baier (1985), p.37.

[27] Baker and Hacker (1984b), p.349.

[28] For suggestive discussion of Wittgenstein's views about such cases see Szabados (1981a) and (1981b).

[29] It is a charge that he lays against Strawson, Rorty, Alston and Shoemaker, for a variety of reasons that I will not consider here.

know how to speak a language just is for me to be able, among other things, to express my beliefs and wishes. If I cannot do this, then I am not a capable speaker of the language. Since I cannot be totally deluded about the meanings of my words, it is a consequence of the fact that I am a language-user that I know, by and large, what I believe. A parallel consequence is that I can also be mistaken about my hopes and beliefs in particular cases. The fact of my being a capable language-user does not entail that I do not make mistakes in applying my words. In fact, linguistic understanding presupposes the possibility of misunderstanding, for linguistic meaning is normative--there are correct and incorrect applications of words and expressions, and if there were not, then a "word" could "mean" anything I wanted it to "mean." But if I can make linguistic errors, then those errors can include misapplications of terms like 'believe' and 'hope' and 'expect'. To suppose otherwise is to suppose that there are "magic words" that cannot be misapplied by a capable speaker. So, I can also be mistaken about my intentional attitudes. Yet, such errors cannot be the rule; we cannot always be mistaken. The shared understanding that characterizes anything complex enough to be called a "language"[30] requires a recognizable degree of regularity of usage.

But why do I not know another's mind with the same reliability that I do my own? An example will help us with Davidson's answer:

> The assumptions are just these: you and I both know that I held the sentence 'Wagner died happy' to be a true sentence when I uttered it; and that I knew what that sentence meant on the occasion of its utterance. And now there is this difference between us, which is what was to be explained: on these assumptions, I know what I believe, while you may not. (FPA 110)

According to Davidson, to understand what belief I express when I say "Wagner died happy" you must devise a hypothesis and check it against the available evidence. That you might not know what I believe in this case stems from the fact that the process of interpretation "cannot be the same for the utterer and for his hearers" (FPA 110). He continues:

> A hearer interprets (normally without thought or pause) on the basis of many clues: the actions and other words of the speaker, what he assumes about the education, birthplace, wit, and profession of the speaker, the relation of the speaker to objects near and far, and so forth. The speaker, though he must bear many of these things in mind when he speaks, since it is up to him to try to be understood, cannot wonder whether he generally means what he says. (FPA 110)

This does not mean that I cannot know what you are thinking, that I am ignorant of your hopes and ambitions, your beliefs and desires. There

---

[30] The fact that it cannot be said just what degree of regularity is required is no objection here (contrary to a suggestion made in conversation by Akeel Bilgrami). I don't know how many grains of sand it takes to make a heap, but I know one when I see one. "Shared understanding" should, by the way, be taken with the caution I advised in discussing the idea of a culture in Chapter 4.

is no serious *philosophical* threat to my knowledge of others, but I can misunderstand you, and you can conceal your attitudes from me. The point is simply that there can be an intelligible asymmetry between first-person claims and second- and third-person claims.

Davidson captures an important difference between the roles of speaker and listener, but it might be tempting to accuse him of the very fault he finds with others. Davidson's criticism of philosophers who agree that there is an asymmetry between first-person knowledge claims and second- or third-person knowledge claims was that they tend merely to restate the asymmetry, rather than explain it. Might one not make this same complaint against the suggestion that speakers stand in another relation to the process of interpretation than do listeners?

This complaint misses the point, for Davidson wanted to explain the differing degrees of epistemic *authority* that attach to these various knowledge-claims, and his suggestion does give an explanation: my relation to my own linguistic abilities differs from your relation to them and from my relation to your linguistic abilities. From my perspective, for *me* to be a capable speaker is generally for me to know my own mind, but from that same perspective, for *you* to be a capable speaker is for your utterances and accompanying behaviour to be interpretable by me, not for me to be authoritative about them.

Even if one accepts Davidson's position as an explanation, and not just a restatement of the special authority that can adhere to first-person judgments, one might press another worry: maybe it is true that for me to be a capable speaker I must know my own mind, by and large, but this takes *self-knowledge* for granted. Isn't it the asymmetry between self-knowledge and knowledge of others that wants explaining, and isn't the first-person authority of *utterances* parasitic on this asymmetry? Davidson offers us a quasi-transcendental constraint on the possibility of speaking a language--if I am to be able to use a language, then I must know my own mind and better than I know yours. But granted that I do know my own mind better, how is this possible?

The proper response to this worry, I think, is to observe that self-knowledge is not only necessary for my linguistic abilities, but also sufficient. To be said intelligibly to know what I believe, desire, hope, fear etc., I must be able to put it into words. Self-knowledge, is a kind of know-how; it is the ability to articulate what I think and feel. If I cannot say or indicate what I believe and desire, then I cannot know what I believe and desire. For to be said properly to know what I believe and desire, it must be possible that I be mistaken in my second-order beliefs. But if I cannot express what I believe and desire, then I cannot very well be wrong or right about it--though others might be. So, I have neither knowledge nor belief concerning my first-order beliefs and desires in such an instance.

V. Asymmetry: Shared Experience and Trust

I have supported Davidson's account of the asymmetry between self- and other-knowledge. Without revoking that support, I shall now suggest that this asymmetry is a matter of degree and can disappear altogether in some cases. Sometimes I know other minds as well, and in the same way, as I know my own, and sometimes I know my own mind as poorly, and in the same way, as I know the minds of others.

There is one obvious respect in which the asymmetry between self- and other-knowledge is not absolute. I *can* be mistaken about my own beliefs and desires. Moreover, I can deceive myself--not quite by

lying to myself, but by engaging in bad faith[31]--and I can be led by ideology or psychological disorder to give poor articulations of my attitudes, which in subtle ways frustrate or do not satisfy fully my inchoate hopes and desires.[32] I can also hold things true of myself that are belied by other aspects of my behaviour. In such cases self-*interpretation*, often with the help of others, may be what I need to obtain self-knowledge. I shall say more about such cases in Chapter 8.

But there are other respects in which knowledge of self and others can run parallel. Consider the case in which I formulate my belief that the milk is in the fridge by saying, "The milk is in the fridge." Typically, when I say this, I know what the words mean, and so, what I believe. I have no need of interpretation in order to understand myself here. But *neither do you*. Provided you understand English and do not come from a part of the world where my accent is too unusual, or there are no refrigerators, or they are not called "fridges," you typically know what I believe making this utterance, and you do not know it by *interpreting* my behaviour, because I behave in just the way that you expect me to. It's not that you don't have access to my non-linguistic behaviour; rather, you don't need it. I might be in another room, when you shout to me, "Where's the milk?" And I respond as loudly, "It's in the fridge!" To be sure, you *can* be mistaken about what I believe when I say this, but so can I, and there are cases in which neither of us is wrong and in which it would be as strange for you to raise doubts as it would be for me. If I respond to your question by pointing at the grocery-bag by the doorway as I say, "It's in the fridge," then you must resort to interpretation. But it doesn't follow that your understanding of me is always like this, nor that our understanding of others *could* always be like this. For if you know that I can be mistaken about myself, or that you can be mistaken about me, it is because you have *learned* that there are exceptions to the frequent tendency to get things right. I can no more learn a language if, in the process of learning, I usually get things wrong about others than I can learn a language in which everyone always lies to me. If lying were the standard case, if everyone lied all or most of the time, it would not be lying--those words and behaviours would have a different general use and a different meaning.[33]

These observations highlight an important factor in determining the degree of asymmetry present between first-person and other-person knowledge-claims--*viz.*, the degree of experience and background-knowledge shared by speaker and auditor. There is an obvious level on

---

[31] Self-deception cannot simply be lying to oneself, as Sartre points out, because that would require fixing one's attention on the very truth being denied. See Sartre (1956), p.87f. What makes self-deception possible, he proposes, is the ambiguous articulation of the truth to be avoided. See Hymers (1989a). Davidson also points to the difference between lying and self-deception, though without any reference to Sartre. He argues that self-deception depends on the possibility of "partitioning" one's mind in some way, but does not offer any account of the mechanisms that achieve this. See Davidson (1985).

[32] That beliefs etc. can be inchoate will be argued in Chapter 8.

[33] If I were the only one to whom everyone always lied, then I would be deluded; I would speak a contingently private language.

which a shared background is relevant to the question of asymmetry. That is the level of language. If we do not share a language, you and I must often resort to interpretation of each other's behaviour to know each other's beliefs and desires. If we share a language, then the need for interpretation is automatically lessened. If we share, not just a language, but a dialect, then the room for misunderstanding undergoes a further decrease. And in such a case we will probably find that we share a repertoire of meaningful gestures, mythology, folk-tales and traditions, and so on. We share a culture, or even a sub-culture. As we move to the level of colleagues, family-members, friends and lovers, the amount of shared experience tends to increase and so, often, does the degree of our transparency to one another. Here a knowing glance may be all that is needed to convey the thoughts of another. These are circumstances under which I sometimes know what another will say before she says it, or under which only I recognize a casual remark as a joke or an insult or a request for attention. But no matter how well I know a person, I am not infallible in judging her thoughts and feelings. Indeed, I am not infallible in judging (when I bother to do so) my own intentional attitudes.

What prompts Davidson to think that other-knowledge always depends on interpretation is, I think, a lack of attention to the importance of shared experience. And this lack can plausibly be traced to his emphasis on the scenario of radical translation. Indeed, we might see his rejection of the idea of a conceptual scheme as contributing to his overemphasis of interpretation. In focusing on the need to attribute what we take to be true beliefs to speakers of a language under radical translation Davidson overstates the degree of uniformity across the pragmatically delineated conceptual schemes that I have argued we should admit into our analysis. He is thus led to apply the same criteria to all cases of understanding others, and the criteria that he inherits from the Quinean scene of translation favour the paradigm of interpretation. By focusing on practical differences, or their absence, my view discourages this temptation.

Now, I said above that I can no more learn a language in which I misunderstand others *as a rule* than I can learn one in which everyone always lies to me. This mention of lying suggests a role for trust in cases of other-knowledge that resemble cases of self-knowledge by being rooted in our capacities as language-users. When I tell you that the milk is in the refrigerator, I may be lying to you, and if I do it well, then you will be wrong in assuming that you know what I believe. If, however, your trust in me is well-founded and it is not a case in which you have reason to think me confused, then you will be justified in taking yourself to know what I mean. You can be justified and still be wrong in particular cases, but you cannot be wrong always, because in general, trust is not merely an option for human beings. Trusting is something that, as Annette Baier has written, we seldom *decide* on, but *find* ourselves doing. "We inhabit a climate of trust as we inhabit an atmosphere and notice it as we notice air, only when it becomes scarce or polluted" (TA 234).

Before these remarks can engender confidence we need to examine the role of trust more carefully. Although my fallibility in judging the intentional phenomena of another is a logical consequence of the objectivity of her intentional phenomena, I can make mistakes here for the further reason that here I am most vulnerable to deception. Trust is relevant to the asymmetry between self- and other-knowledge, I shall argue, in the following way: trust is justifiable if and only if

its preservation does not require either self-ignorance or ignorance of the trusted party's intentional phenomena. Thus, if my automatic, non-reflective assessments of another are to be reasonably accurate, I must be in a relation of justifiable trust with her. Let us see why.

One striking feature of trust, Baier observes, is its pervasiveness. Scarcely anything that matters to anyone can survive and grow in the absence of trust. Trust, says Baier, is a "phenomenon we are so familiar with that we scarcely notice its presence and its variety ..." (TA 233). Both qualities are confirmed by the many sorts of trust-relationships in which we find ourselves--relations with, not just the most, but also the least familiar of people, and "even with declared enemies," whom we trust "not to fire at us when we lay down our arms and put out a white flag" (TA 234). Baier continues:

> We do in fact, wisely or stupidly, virtuously or viciously, show trust in a great variety of forms, and manifest a great variety of versions of trustworthiness, both with intimates and strangers. We trust those we encounter in lonely library stacks to be searching for books, not victims. We sometimes let ourselves fall asleep on trains or planes, trusting neighbouring strangers not to take advantage of our defencelessness. We put our bodily safety into the hands of pilots, drivers, doctors, with scarcely any sense of recklessness. We used not to suspect that the food we buy might be deliberately poisoned, and we used to trust our children to day-care centres. (TA 234)

We depend on others as we try to fulfil our plans and projects. But it is a special dependence, a reliance on the "good will" (TA 234) of others. It is not the kind of dependence that I have on the clock-radio that wakes me every morning; it is more like the dependence that I have on the good intentions of the clock-radio's maker. I trust her not to make a clock whose alarm sounds an hour later each day. I can rely on a person without trusting her. I can rely on her to do what she can to frustrate my plans, and I can rely on regularities in her character without relying on her good will:

> Kant's neighbours who counted on his regular habits as a clock for their own less automatically regular ones might be disappointed with him if he slept in one day, but [would] not [have been] let down by him ... (TA 235)

So, we find ourselves depending on the good will of others, wisely or unwisely. And we do "find ourselves" in these situations. Trusting is often not something that we choose to do, but something we find ourselves doing, and only then do we make a decision about whether or not to continue trusting. We are often caught up in the game of trust without being asked to play and without having a special end that will, we think, be furthered by entering a trust-relationship. "The ultimate point of what we are doing when we trust may be the last thing we come to realize" (TA 236).

But with trust comes the vulnerability to which I alluded earlier. When I trust someone, I put myself and the things with which I entrust her in a position such that I and they can be hurt. I can trust a friend with a book that I lend to her. I expect her to return it in more or less the same condition it was in when I lent it. But in

trusting her to do so, I make myself and that with which I entrust her vulnerable. She can, if she chooses, betray my trust by flinging my book into the North Saskatchewan River. If I trust her with my safety, as I do bus-drivers and (some) motorists, she can betray my trust by driving off a bridge or trying to run me over.

> Since the things we typically do value include such things as we cannot singlehandedly either create or sustain ... we must allow many other people to get into positions where they can, if they choose, injure what we care about, since those are the same positions that they must be in in order to help us take care of what we care about. (TA 236)

How is this vulnerability relevant to the asymmetry between self-knowledge and knowledge of others? That asymmetry consists, recall, in the superior justification for self-knowledge claims--superior, that is, to the justification for claims about others. I contend that the asymmetry varies according to the degree of justifiable trust that obtains between or among the parties involved. Other things being equal, in a relation of justifiable trust the asymmetry between first-person authority and second- or third-person authority will be lower than in a relation without justifiable trust, and in a way that does not decrease first-person authority. Why should this be so?

The reason is clear. If I am in a trust-relationship with another, then typically I count on certain things being true of her. I take myself to know something of her intentions, especially with respect to me and that with which I have entrusted her. But if believing those things that I take myself to know about her intentional phenomena cannot be justified, my trust is unjustifiable. In such circumstances I can discover that I do not know a person as well as I thought I knew her. If I trust a friend to water my houseplants while I travel, and she uses the chance to create an infestation of parasites, then I am deluded, and my judgments about her thoughts and feelings lack the authority of my judgments about my own thoughts and feelings. So, is knowing another's intentions regarding myself a sign of a justifiable trust-relation? To clarify how the asymmetry between self- and other-knowledge varies we need an account of justifiable trust.

Baier proposes a test for "the moral decency" (TA 255) of trust:

> [I]ts continuation need not rely on successful threats held over the trusted, or on her successful cover-up of breaches of trust. (TA 255)

Her idea is the intuitively plausible one that there is something morally rotten about a trust-relationship, if it would not survive the truster's confrontation with "the facts." Suppose I borrow a book from you and do not return it. When you confront me, I tell you that I have forgotten to return it, but that I will do so the next day. This may be enough for you to continue trusting me, and if the next day I claim to have forgotten again, you may yet be inclined to believe me and go on trusting. But if time passes, and I continue to make excuses, your confidence will be undermined. "He is lying," you might tell yourself, or "If he's not lying, he's certainly not treating my trust with much respect." Your loss of trust, we could say, is *justifiable.* It would be reasonable for you to deny my next request to borrow something.

In this example you have learned that your trust in me depended on my deceiving you. And this awareness makes manifest that your trust in me is ill-founded. Baier paraphrases her point: "A trust relationship is morally bad to the extent that either party relies on qualities in the other which would be weakened by the knowledge that the other relies on them" (TA 256). An unjustifiable trust-relationship relies on a certain kind of ignorance about the trusted on the part of the truster. The truster does not know that which she assumes to be true, or would regard to be true if she bothered to think about her trust. (Trust can be "unconscious" (TA 235).) But by the same token, if her trust is justifiable, she *can* know--if she reflects on the point--that which she presupposes about the trusted's intentions toward her and toward that with which she entrusts the trusted. So, if your trust in me is justifiable, then you will normally know, when I say that I have forgotten your book, that I speak the truth and that I have certain intentions regarding the treatment of your property, its prompt return and you as the person whose trust I recognize and wish to preserve.

But Baier's test does not cover certain cases that are of special interest in a discussion of self-knowledge. Baier supposes that upon realizing that my trust in another has depended on my ignorance of that other's intentions toward me, my trust will tend to be weakened. But there is no guarantee that my discovery of my ignorance of the trusted *will* undermine my trust. Upon learning that my friend has deliberately introduced parasites into the soil of my houseplants I might continue to trust her, because I believe--wrongly, let us suppose--that her actions were justified, that I have done something to deserve this. Maybe I have a poor self-image. I view myself as less worthy than the person whom I have trusted, because my society fosters negative attitudes toward people of my gender, sexual preference, cultural background, socioeconomic status, etc. Or perhaps, in the absence of such factors, I am so shaken by the revelation that my trust has been abused that I slip into self-deception to avoid recognizing her actions for what they are and continue to trust her.

These cases are ones in which my trust is *not* justifiable, but in which Baier's test fails to reveal this lack of justifiability. They are not, however, cases of which Baier is unaware. She writes:

> I have used the phrase "tend to destroy" in the test for moral decency in the assumption that there is a normal psychology to be discerned and that it does include a strong element of Platonic *thumos*. Should that be false, then all sorts of horrendous forms of trust may pass my test. (TA 257)

The examples offered above are examples of trusting individuals whose trust depends on their possessing an "abnormal psychology." How might Baier's test be supplemented to cover these abnormal cases?

The cases that I have raised stress varieties of self-*ignorance*. In my self-deception I have grasped a truth that could normally undermine my confidence in my friend, but I have hidden the truth from myself. I have nurtured an ignorance about my own beliefs. Similarly, if I am among the ranks of the oppressed, I may be denied, in the words of W. E. B. DuBois, "true self-consciousness."[34] I may wrongly judge myself

---

[34] Quoted in West (1989), p.142.

in the dismissive, subordinating terms of my oppressors, and suppose myself unworthy of having my trust respected or returned.

I have already interpreted Baier's test for the justifiability of trust as emphasizing my knowledge, or lack thereof, of others. I now propose the following extension of that test:

> Not only does the continuation of a justifiable trust-relation not rely on ignorance of the intentions of the trusted, but neither does it rely on the truster's *self-ignorance*.

This extension says something of the "normal psychology" to which Baier adverts in limiting the applicability of her test. A person with a normal psychology is not ignorant of herself in a way that impedes her ability to know the relevant intentions of the trusted or to reason or act on that knowledge. However, my extension of Baier's test is not meant to give a rigorous account of "normal psychology." I wanted to show that the asymmetry between first- and other-person authority varies in relation to the degree of justifiable trust as follows: to the extent that trust is justifiable, *ceteris paribus*, the asymmetry between the truster's knowledge of herself and her knowledge of the trusted decreases. This has, I think, been made credible.[35]

It is tempting to take a further step: if I have self-knowledge, it is because I can speak a language, and if I can speak a language, it is because I have learned it from others, and if I have learned it from others, it is because I have not been thoroughly deceived in relying on their good will. Thus, I can know my own mind, *because* I can know the minds of others, and I know my own mind--in the first instance--in the same way as I know the minds of others. Persons are, to borrow again from Baier, "second persons."[36] This is not simply to repudiate the asymmetry that Davidson finds obvious. Nor is it to accept the Rylean doctrine that self-knowledge is like knowledge of others in depending on inference and interpretation. But the asymmetry is a matter of degree, as Ryle thought, and it emerges only after I have begun to acquire the requisite linguistic skills and the relative independence that come with growing out of infancy.

It should not be protested that the proposed connection between self-knowledge and knowledge of others rests on an argument which, as

---

[35] A decrease in the asymmetry between self- and other-knowledge claims does not guarantee the justifiability of a trust-relation. I might trust another, because my self-ignorance makes me ignorant about her. Here the asymmetry decreases, but for the wrong reasons.

[36] Baier (1985), p.90. Baier actually means to make a distinct, albeit related point, with this expression. Her claim is that I come to self-consciousness in learning to identify myself with others' uses of the second-person pronoun: "My first concept of myself is as the referent of 'you,' spoken by someone whom I will address as 'you'" (*ibid.*, p.90). But the points are connected, I think. Lorraine Code proposes an adaptation of Baier's account of second persons, which emphasizes the importance of interpretation in other-knowledge, but she is concerned primarily with knowledge of others in the context of the analysis of oppression. I shall take up some related concerns below. See Code (1991), pp.27-109. See also Burns (1989).

Quine might say, has the form of a closed curve in space. This is a tempting supposition, since I have claimed that trust is linked with self- and other-knowledge, which in turn require language, but also that language-learning requires that my dependence on the good will of others not be systematically betrayed. (To repeat: a language in which everyone always lied to each other would not be a learnable language.)

Notice, however, that my central claim is that, *ceteris paribus*, the asymmetry between self- and other-knowledge is decreased in a relation of justifiable trust. And this is compatible with supposing that one party is, for now, in a general state of ignorance regarding herself and the intentional phenomena of others, because she has not yet learned a language. Justifiable trust does not require linguistic competence on the part of both parties; nor does it require that they both possess knowledge of themselves and of each other. What makes a trust-relationship justifiable is that its continuation does not *depend* on the ignorance of the truster.

This is compatible with saying, for example, that the child who is just beginning to learn her first language can have justifiable trust in her caregivers. She grows into relations of trust with others, as she learns her language. At the start she is no more capable of trusting in the *articulated* way that a language-user is than she is of speaking a language--both remain potentialities for her. Nonetheless, she is, as an infant, still dependent on the good will of others, and while she cannot yet assess for herself the justifiability of her trust, she also acquires that capacity for assessment, as she learns to speak and to understand. It is, thus, possible for her in time to learn her own intentional phenomena and those of others whom she trusts; so, the test for justifiable trust is applicable, albeit after the initial trust-relation arises. (There is nothing very unusual in this temporal delay--it may not be practically possible, or even desirable, to test a new trust-relationship.)[37]

So, there is no circularity involved in my claims, on the one hand, that knowledge of self and others depends first on language and, in turn, on justifiable trust, and on the other hand, that the truster will typically know without interpretation the intentional phenomena of another and of herself, if her trust in that other is justifiable. Without some minimal resolve or tendency on the part of others not to

---

[37] See TA 260. This does not mean that a pre-linguistic child has, *e.g.*, beliefs about other persons' intentional phenomena (although such a child can have certain kinds of beliefs, as I shall argue in Chapter 8). One can rely on the good will of another without knowing what it is on which one relies. As Baier observes, "One constraint on an account of trust which postulates infant trust as its essential seed is that it not make essential to trusting the use of concepts or abilities which a child cannot be reasonable believed to possess" (TA 244). So, we might conjecture that a non-human animal can trust, too--but its trust is neither justifiable nor unjustifiable in the manner that Baier has in mind. Her test requires that trust be able to endure the recognition of the mechanisms of its genesis and preservation by those involved in the trust-relation. But we might consider extending the notion of justifiable trust to trust-relations with animals by imagining ourselves giving or rejecting justifications for similar trust in our own case.

abuse my dependence on their good will I would not learn a language, and I could muster little in the way of knowledge of self or others.

I have been examining the way in which knowledge-claims about others can carry a justification comparable to that characteristic of first-person authority. But there is another respect in which the asymmetry of self- and other-knowledge can be eroded. There are cases in which the asymmetry in question can be affected by my own self-ignorance--cases in which I do not know my own mind because, *e.g.*, I am wrestling with a moral or prudential dilemma, or I am self-deceived, or I am suffering from ideological delusion, or I have a real neurosis, rather than a merely epistemic one. Cases of this sort emerged above in my consideration of Baier's test for justifiable trust. What also emerged was a contrast between articulated and unarticulated intentional phenomena, as exemplified by the possibility of infant-trust. But are "unarticulated intentional phenomena" intentional phenomena at all? Need one be a language-user to have beliefs and desires?

I shall consider these issues in the next chapter, where the topic of self-ignorance will also lead to an examination of the idea of self-unity, which is sometimes thought to be undermined by the possibility of self-ignorance. In accounts of the self, I shall suggest, as in accounts of self-knowledge, the dichotomy of the subjective and the objective is often at work. I will try to make sense of the possibility of self-ignorance and self-division in a way that does not rule out either the possibility or the desirability of some kind of self-integration.

Chapter 8: Self-Interpretation and Self-Unity

As we have seen, Davidson grants that the asymmetry between self- and other-knowledge can be decreased or eliminated by self-ignorance. If I am self-deceived, or faced with a difficult moral decision, or deluded by ideology or mental disorder, my normal linguistic faculties may not suffice for self-knowledge. I may need the help of self-interpretation and, maybe, the aid of others in consciousness-raising or therapy. The asymmetry between self- and other-knowledge decreases here, since the former resembles those cases of the latter that Davidson holds paradigmatic. But a story of self-ignorance requires a story of the unarticulated intentional phenomena of which one can be ignorant. Such a story was also assumed in Chapter 7 in my remarks on trust. So, I shall begin by examining a related question: can non-linguistic animals have intentional attitudes?

Having argued that unarticulated intentional phenomena are possible, I shall turn to Charles Taylor's views on self-interpretation. His position can augment Davidson's by explicating the articulation of hitherto unarticulated beliefs and desires. It also reveals the kernel of truth wrapped in Kant's talk of self-affection and inner sense, which we met in Chapter 6. In many cases the new articulation of intentional attitudes entails an imposition of form on what had been inchoate, and more than one articulation can shape the same inchoate drives and attitudes. As well, which articulations one has available can affect one's "field of possibilities," so that self-knowledge as self-interpretation can matter to one's capacity for agency.

But the very possibility of self-ignorance affects our conception of the self, too. My critique of Burge's views in Chapter 7 was designed to show that the dualism of the subjective and the objective affects contemporary accounts of self-knowledge. In this chapter I will argue that this dualism also influences some contemporary accounts of the self. This is clear in the discourse of recent radical theory. The theme of my discussion will be self-unity with a focus on two issues: first, is a unified self an attainable goal--even as an *ideal*? And secondly, is unity of the self as an ideal a *desirable* goal?

The former issue arises from rejecting the incorrigibility of self-knowledge and recognizing the possibility of self-ignorance. It has been pursued vigorously by European thinkers and Anglophone literary theorists in the second half of our century. The latter issue arises partly as a result of the former, but also as a result of room that has been made for, and achieved by, voices from the "borderlands," most strikingly in the work of many feminist writers who have begun to confront the untidiness of cultural and personal identity. I shall argue that abandoning the Cartesian "unified subject" does not entail abandoning self-unity. Indeed, the very possibility of recognizing oneself as "plural" or "fragmented" depends on a kind of grammatical unity of the subject of one's self-descriptions. That there is such a thing as language at all implies that there are selves--grammatical unities like "I" and "you"--for whom language can have meaning. The question of whether self-unity is desirable, then, is a question about whether a particular array of self-descriptions is desirable, and this question must be asked in particular cases.

I. Unarticulated Beliefs: Human Beings and Other Animals

I said with Kant in Chapter 6 that the succession of my conscious states was insufficient for my knowledge of my states as determined in time; such knowledge requires a "complex representation" of my states *as in some determinate order*. We can now see that talk of complex representations should be replaced by talk of linguistic capacities. Other than knowing of spatio-temporal objects, what enables me to know that my "states" are determinately ordered is that *I can say so*.

The parallel with Kant is worth extending, for it helps to clarify how, if I did not have linguistic abilities and could not express my beliefs and desires, I could yet have beliefs and desires. This in turn clarifies what it means to have linguistic capacities and to have unarticulated intentional phenomena. The complex representation that Kant held necessary for self-knowledge he also regarded as a matter of the self's *affecting* the self. In forming a complex representation of my states I make something true of myself that was not true before: that I have this complex representation. In doing so, I impose a certain order on the "chaos" of my sensory input.

But to say this is misleading in three respects. First, it is in learning a language that we acquire self-knowledge, and it is better to stick to linguistic terms and leave mental "representations" aside. Secondly, since we inhabit a more or less regular world of organisms and inanimate spatio-temporal objects, it is a distortion to think of our sensory input as chaotic. We biological organisms are subject to whatever law-like regularities there are in the universe, and we could not exist without them, let alone come to self-consciousness by a fortuitous chance of evolution. Thirdly, the sense in which I know myself as *affected* by myself is not just the trivial one in which I add another "representation" to my repertoire. Saying *what* I think and feel can affect the character of those thoughts and feelings, taking them from the realm of unspoken impulses to that of articulated beliefs and desires, making clear to me what had been obscure.

But what is it that is clarified here? Am I articulating a belief or a desire that I already had? Should we say that unarticulated beliefs and desires are "beliefs" and "desires" at all? I said in Chapter 7 that trust could also be unarticulated, and this point was important to my account of the graduated nature of the asymmetry of self- and other-knowledge. Let me deal with these issues by examining a related problem of some intrinsic interest: should we attribute intentional phenomena to non-human animals--specifically, to animals incapable of language-use? I shall focus on beliefs for the sake of (relative) simplicity. Whether my results apply to other intentional attitudes I do not wish to prejudge, but if it is plausible to attribute beliefs--those most cognitive of attitudes--to wordless creatures, then it will also be plausible to attribute other intentional phenomena to them.

In answering our question we must not forget the normative character of truth, for if beliefs can be true or false, then they fall into the scope of the normative. We must especially heed the internal link between truth and warrant. If a belief or a sentence is true, then it *could be* justified.[1] But what sense of "could" is relevant here?

---

[1] See Chapter 2, §§V-VI.

A partial answer is that if a sentence or belief is true, then it is really possible to justify it. But for whom is it so possible? It is tempting to respond that a true sentence or belief is one that could really be justified by the one who utters or holds it. If so, then non-linguistic animals do not have beliefs, since they are in no position either to articulate beliefs or to justify them.

But this is to confuse criteria of knowledge with criteria of truth. For a range of cases quite central to the concept of knowledge an epistemic agent must be able to justify her belief, if it is to be agreed that she has knowledge. But in order to utter a truth, or in order to have a true belief, I need not be able (at least in many cases) to give a justification for holding it true. Still, in holding a belief true, we commit ourselves to the real possibility of its being justified by a competent language-user.

These remarks suggest that it is consistent to regard truth as an epistemic notion--as I think we should--and to hold that non-human animals have beliefs.[2] They cannot articulate or justify those beliefs (if they have any) themselves, but this need not prevent us language-users from doing so. Is this to say that we simply "project" beliefs and desires onto non-human creatures? What if there were no language-users? What about "possible worlds" in which no language-users exist, but in which there are, say, dogs and cats? It is surely a logical, physical and evolutionary-biological possibility that language-users might not have evolved. Does my position entail that, had this been the case, dogs and cats would have had no beliefs? This result is unwelcome, for it suggests that non-linguistic animals do not really have beliefs, even if it is convenient or comforting for us to pretend that they do. If animals do not believe anything apart from our so describing them, then their "beliefs" are nothing but our projections.

However, no such consequence follows from my account of truth. What makes truth normative is the fact that there are no truths that could not really be justified or expressed. Now, were there no language-users, then no truths would be *expressed* or *justified*, but they might still be *expressible* or *justifiable*. A world in which there are only non-linguistic animals is not a really possible world, because the really possible worlds are worlds in which there are language-users.[3] But *at such a world* it is still really possible that the beliefs of wordless creatures should be expressed, because there is another world much like that one, but at which language-users do exist.[4] We must not

---

[2] Thanks to Istvan Berkeley, Sarah Hoffman and Joachim Ludwig for helping me clarify my cloudy thoughts on these issues.

[3] See Chapter 2, §II.

[4] In other words there is a really possible world (at which there are language-users), *accessible* to the really impossible world at which there are no language-users. But there is no really possible world similarly accessible to a merely logically possible world at which we have always been brains in a vat. Such a world would have to be one in which we had always been brains in a vat, but in which there were other language-users "outside" the vat--the evil-demon hypothesis is similar to such a world. However, the mere presence of language-users outside the vat would not change the fact that we could express neither our "envatment," nor the fact that we could not express our

confuse the unexpressed with the inexpressible. The normativity of truth, properly understood, in no way conflicts with ascribing intentional attitudes to speechless animals. The only question is whether one can have beliefs without being a language-user. Can one?

Speakers of English and of other languages often talk of non-human animals as if they have intentional attitudes. If I put my hand to my mouth and pretend to be eating, a dog watching me will often display the kind of behaviour that it would display if I were really eating. The dog, we might say, *believes* that I am eating. But why think this?

Richard Rorty suggests that in such cases we extend our ascriptions of belief to dogs and others among the "better-looking animals" as a courtesy that we show toward "potential or imagined fellow-speakers of our language ..." (PMN 190).[5] Indeed, we extend the same courtesy to pre-linguistic children, according to Rorty.[6] But I think that Rorty underrates the importance of non-linguistic conduct in our ascriptions of intentional phenomena to language-users, and so, he also neglects the importance of non-linguistic conduct in our ascriptions of intentional phenomena to animals incapable of linguistic behaviour.

Consider something that Wittgenstein says about the intentional phenomena of animals in Part II of *Philosophical Investigations*:

> One can imagine an animal angry, frightened, unhappy, happy, startled. But hopeful? And why not?
>
> A dog believes his master is at the door. But can he also believe his master will come the day after tomorrow?--And *what* can he not do here?--How do I do it?--How am I supposed to answer this?
>
> Can only those hope who can talk? Only those who have mastered the use of language. That is to say, the phenomena of hope are modes of this complicated form of life. (PI II 174)

Wittgenstein's point might be this: certain sorts of intentional phenomena do not require the presence or believed presence of their

---

[4] envatment. That is, as we saw in Chapter 2, not only could we not say with justification, "We are brains in a vat," but we could not say with justification, "We could not say with justification that we are brains in a vat." Nor is it *true* at a world without language-users *that it is a world without language-users.* "Dumb brutes" have no beliefs about "language" or "language-users," and only their beliefs count as unexpressed truths "at" such a world. So, neither is there any conflict between my saying, on the one hand, that the unexpressed beliefs of animals in a world without words would still be expressible and my saying, on the other hand, that "If language-users had not evolved, there would still have been a world, but there would not have been any *truths* about the world" (Putnam, 1992b, p.433), provided we build into the "If"-clause that animals capable of normative behaviour had not evolved either. See Chapter 2, §VI.

[5] For a helpful discussion of Rorty's position see Hunter (1982), pp.634-638. Rorty deals with knowledge and conceptual abilities, but these issues are intertwined with the attribution of beliefs.

[6] See PMN 190.

objects. I can hope for--only for--what I believe I do not have, and I can express hope, only if I can speak. No non-linguistic behaviour will suffice to express my hope (without a complex context that relies on the use of language).[7] The dog, similarly, would have to be able to say that it expected some human being to show up the day after tomorrow, in order to have such a belief or an expectation.

But does this rule out the possibility of the dog's believing that a particular person is at the door? Isn't the sound of the key in the lock or the familiar footfall on the stair a present indicator that needs nothing, no word, to stand in for its association with a particular person? Here we might say of a human being or a dog that it believed a specific person to be at the door, and our reasons for the attribution would be quite similar--the turning of the head, a certain facial expression or raising of the ears, a movement toward the door. (A human being, unlike a dog, could suppress these signs.) Beliefs are not all cut from the same cloth. Some I can have only as a language-user--their phenomena "are modes of this complicated form of life" (PI II 174). But this does not rule out pre- or a-linguistic beliefs.[8]

Attributions of intentional attitudes happen against a background of prior or presupposed attributions and, amongst language-users whom we understand, prior articulations. If we suppose a person to believe in the value of exercise, then we expect a certain coherence between this belief and the other beliefs and desires that we take her to have. What she has said in the past will play a role in our appraisal. But whether or not she exercises will also be relevant. It is not only linguistic behaviour that matters here.[9]

As we saw in Chapter 3, §V, a people who seemed to display linguistic behaviour, but whose utterances bore no discernible, practical connection with the rest of their behaviour, would not properly be judged language-users. There would not be, as Wittgenstein says, "enough regularity for us to call it 'language'" (PI §207). Similarly, the Turing-machines of Putnam's test for reference, considered in Chapter 2, §I, would not properly be taken to use language, if their imitations of conversation did not have some regular connection with the world. A computer with no "sensory transducers" and no means of manipulating or interacting with the world would not be engaging in *linguistic* behaviour. Meaning and

---

[7] Is this right? When I finish eating or pretending to eat, doesn't the dog look at me *hopefully*? Are there two sorts of hope, linguistic and non-linguistic? Consider Wittgenstein's ambivalence:

> Compare the expression of fear and hope with that of 'belief' that such-and-such will happen. -- That is why hope and fear are counted among the emotions; belief (or believing) however, is not. (Wittgenstein, 1980, §596)

[8] We should bear in mind Wittgenstein's remark that "If a lion could talk, we could not understand him" (PI II 223). This seems to clash with attributing beliefs to the lion, since we could justifiably do so only if it could in principle be understood. But it is possible that Wittgenstein has in mind a practical incommensurability (of the sort sketched in Chapter 3) of human and leonine conceptual schemes.

[9] For this point see Malpas (1992), pp.83-86.

reference are possible only where there is practical activity. Only because our utterances have roles to play in the pursuit of our interests do they have meaning at all.

Now, if we link this observation to the holistic nature of belief-attribution, we see that for language-users ascriptions of intentional phenomena rest, not just on linguistic behaviour, but on the practical activity crucial to the meaningfulness of that behaviour. To interpret the people of Wittgenstein's example as language-users, we would need to be able to find regular connections between what they said and what they did. But it does not follow from this that we cannot interpret their behaviour at all, and so, it does not follow that we cannot meaningfully attribute beliefs and desires to them.

Much the same holds true for actual wordless animals. Dogs and cats and even chimpanzees do not display anything like the sophistication of the people of Wittgenstein's thought-experiment, who "carr[y] on the usual human activities" (PI §207). But their behaviour is not simply random, either. Their behaviour can be complicated and orderly and often bears a resemblance to certain kinds of human behaviour. They can display behaviour geared toward the satisfaction of their animal-interests, behaviour that can also be "corrected" for human ends by means of training. They can correct their own behaviour and the behaviour of others of their kind (and other kinds). The dog that herds sheep knows how to herd sheep, and its know-how is displayed in its ability to recognize when the sheep are being successfully herded, its ability to see that a sheep is straying from the flock.[10] There are correct and incorrect applications of intentional terms to such behaviour. The dog can believe that a sheep has strayed from the flock. A cat can believe that a mouse is in the mouse-hole--but not that the period of a simple pendulum varies as a function of the square root of its length. Human beings display their grasp of concepts in linguistic and non-linguistic behaviour, and the latter is crucial to the former. Wordless animals display their grasp of simpler concepts in their non-linguistic behaviour, sharing with us some of what is central to the meaningfulness of our linguistic practices.[11]

---

[10] I am not sure if this amounts to "grasping the contrast between truth and error--between true belief and false belief" (ITI 170) as Davidson hints it would have to if we were to be justified in attributing intentions to non-human animals. If not, then so much the worse for his refusal to attribute intentions to wordless animals.

[11] The simplicity of "doggy-concepts" is sometimes mistaken for an indeterminacy of those concepts. Thus Davidson, focusing on the non-linguistic expression of human preferences, writes:

> [A subject] can be taken to have expressed a preference by taking action, by moving directly to achieve an end, rather than by saying what he wants. But this cannot settle the question of what he has chosen. A man who takes an apple rather than a pear when offered both may be expressing a preference for what is on his left rather than the right, what is red rather than yellow, what is seen first or judged more expensive. (ITI 163)

If we can't interpret the subject's speech, Davidson thinks, we can't

The temptation to think that other animals don't have beliefs stems, I believe, from forgetting or ignoring the importance of non-linguistic conduct for the warranted application of intentional terms.[12]

---

[11] warrantedly draw "the fine distinctions we are used to making in the attribution of thoughts" (ITI 164), and so, "our attributions and consequent explanations of actions will be seriously underdetermined in that many alternative systems of attribution, many alternative explanations, will be equally justified by the available data" (ITI 164). The proper response, I think, is to say that the fineness of a "fine distinction" is relative to our interests in drawing it. That doggy-concepts fail to distinguish an apple from a pear, *e.g.*, says no more about the correctness of human ascriptions of those concepts than the fact that different cultures make different distinctions, relative to different interests, says about one culture's ability to understand another. As I argued in Chapter 3, §IV, the idea that it is fitting to apply the notion of the underdetermination of theory by data to the interpretive understanding of members of another culture incorrectly assumes that normativity can be replaced by some naturalistic notion. Similarly, the thesis of the indeterminacy of doggy-concepts *assumes* that non-linguistic creatures have no concepts, no capacity for normative behaviour. (Davidson concedes that his remarks on "fine distinctions" do not "constitute an argument" (ITI 164).) It is at this point that the difference between Davidson's view of conceptual schemes and my own (Chapter 3) comes into focus, for without the concession that other cultures can have different concepts, the idea that other animals might have any concepts is harder to make out.

[12] And if we treat having a belief as like having a "sentence in the head," or as having an attitude toward a *proposition*, then it is easy to see how we might come to suppose that only a language-user could have beliefs. (Davidson writes of "propositional attitudes," but it is not clear that he should be taken literally. See, *e.g.*, ITI 166. Nor does he hold that "thought" can be reduced to language. See ITI 158.) It may also be tempting to suppose that non-human animals can't have beliefs, because they can't "think" about the object of a belief. But as Wittgenstein suggests, I don't have to *think* about the object of an occurrent belief in order to believe something.

> When a cat lies in wait by a mouse-hole--do I assume that it is thinking about the mouse?
>
> When a robber waits for his victim--is it part of this, for him to be thinking of that person? *Must* he be considering this and that as he waits? Compare one who is doing such a thing for the first time, with one who has already done it countless times. (Wittgenstein, 1980, §829)

Davidson, like Burge, speaks of beliefs as "mental states," as if they were like fear or sadness. And his "anomalous monism" endorses the token-identity of brain-events and mental-events. See Davidson (1970). But, as Hacker observes, it is unclear that beliefs fit the model of mental states. "There are indefinitely many things that I believe at a given time, but I am not in indefinitely many mental states at a given time" (Hacker, 1992, p.250). See also White (1972). Indeed, it is heeding only certain cases of believing that tempts us

What of infants, to whom we are alleged by Rorty to extend belief-attributions as a courtesy? This case is complicated by the fact that very young human beings display, at first, behaviour that is less sophisticated than the behaviour of many animals incapable of normative activity. Perhaps Rorty is right to suggest that our attributions of beliefs and the like to infants are at first a kind of courtesy. But the point at which it ceases to be a courtesy is not at all clear, and if non-human animals can correctly be said to have beliefs, then the end of courtesy and the beginning of genuine interpretation will come much earlier than Rorty's focus on language-acquisition would dictate. Nor on the strength of this reasoning does it seem far-fetched to speak of "infant-trust," as I did in Chapter 7.

Let me now return to the issue of whether or not the unarticulated beliefs of a competent language-user are really "beliefs" before their articulation. The possibility of this is implicit in the claim that non-linguistic animals and pre-linguistic children can have beliefs. The beliefs of mute beasts may manifest themselves in behaviour, but they are forever unarticulated. However, being able to articulate one's beliefs does not require one to articulate them all. A great many beliefs go unarticulated. If it makes sense to attribute unarticulated intentional phenomena to non-linguistic animals, then it makes sense to make similar attributions to more talkative humans.

But there is a further complication. Sometimes when I am asked what I believe, I cannot at first say, and I do not at first know. Maybe the problem is a complicated one, a question in philosophy, for example--"What do you think about the problem of universals?" Or maybe I find myself torn between two courses of action that are of moral consequence--"Should I join the Resistance, or remain here and care for my ailing mother?" Or maybe ideology or psychological disorder clouds my capacity to answer the questions, "What do you believe about this?" "What do you hope for?" "What do you want?" "What do you fear?" I feel like saying of such cases that sometimes (though not always)[13] my intentional attitude does not pre-exist my eventual articulation--at least not in the full sense that its articulation later brings to it. This is not to say that I create my beliefs and desires *ex nihilo*. One does not, in Nietzsche's colourful phrase, "pull oneself up into

---

[12] to think of believing as being in a state or engaging in an activity:

> [W]e form a picture of the man who believes the whole time that he is hearing a low rustle. But not of the man who believes in the correctness of the law of gravity. (Wittgenstein, 1980, §597)

In order to believe that there is a mouse in the hole the cat need not be able to attend to some sort of mental image or to "entertain" a proposition. It is enough that it *waits*--enough, that is, against the background of its behaviour on other occasions. Talk of beliefs as though they were a kind of thing itself contributes to the view that a belief is a mental state. See Hacker (1992), p.257.

[13] For example: I might have certain pathological beliefs or desires that I am prevented by some psychological mechanism from recognizing and articulating for myself. But others may well be able to help set me straight in such matters.

existence by the hair, out of the swamps of nothingness."[14] The "raw materials" of my beliefs and desires constrain the range of their possible articulations. But I started this section with some remarks about how my self-articulations can alter the intentional phenomena that they articulate. Let me now return to the point of those remarks.

II. Self-Knowledge and Self-Interpretation

Much the point that I want to extract from my modified version of Davidson's position with the aid of Kant has been made in recent years by Charles Taylor. It is this: although self-knowledge is fallible and corrigible, it is still true that for some kinds of self-description, to a certain extent, saying so makes it so. When I express my beliefs and desires, Taylor suggests, I sometimes articulate something that was previously vague or confused. I arrange my hitherto disarrayed attitudes, bringing into existence something more definite; there is an element here, not just of discovery, but of creation as well:[15]

> Now these articulations are not simply descriptions, if we mean by this characterizations of a fully independent object, that is, an object which is altered neither in what it is, nor in the degree or manner of its evidence to us by the description. In this way my characterization of this table as brown, or this line of mountains as jagged, is a simple description.
>
> On the contrary, articulations are attempts to formulate what is initially inchoate, or confused, or badly formulated. But this kind of formation or reformulation does not leave its object unchanged. To give a certain articulation is to shape our sense of what we desire or what we hold important in a certain way.[16]

Taylor is concerned with more fundamental matters than my knowledge of my belief that there is milk in the refrigerator. I don't agonize over articulating that belief in the way that I may agonize over choosing a career or believing the story of a friend whom I suspect of deceiving me--perhaps because I do not have as many competing alternatives to confront, but certainly because the consequences are not so weighty. Rather, if someone asks me where the milk is, I formulate my belief thus: "It's in the refrigerator." In saying this, I typically know what I mean, and hence, what I believe. My decision about whether to join the Resistance or care for my mother does not come so easily or automatically. But my doubts and difficulties here are not evidence of

---

[14] Nietzsche (1966), §21.

[15] The Kantian heritage of this view also underlies Sartre's distinction between reflective and pre-reflective consciousness. According to Sartre, I do not know myself as an *object*, and attempting to identify myself this way is a slip into bad faith. See Sartre (1956). But Sartre's pre-reflective consciousness is "an immediate, non-cognitive relation of the self to itself" (*ibid.*, p.12), and is closer to Bilgrami's account (see Bilgrami, 1992) than to Taylor's.

[16] Taylor (1985), p.36.

my linguistic deficiency. My problem is not that I don't know what an obligation is, or what it means to "care" for someone or to "join" an organization--though I *can* be confused about these things. What I don't know are things like what I believe I ought to do, what sort of person I am or want to be. My decision brings greater definition to these matters; it helps to determine--logically, not causally--what I believe, what I want, etc. But I still know my beliefs and wants as I have formulated them, and I come to know them in that formulation.

It will seem as though I am back-pedalling here. I had claimed to be supporting Davidson's fallibilist explanation of the asymmetry between self- and other-knowledge (with the *proviso* that the asymmetry is a matter of degree), but am I not giving this all away by conceding that my believing that I believe that *p* makes it true that I believe that *p*? Isn't this Burge's claim that "One is thinking that *p* in the very event of thinking knowledgeably that one is thinking it" (ISK 654)?

My proposal can be elaborated without slipping into subjectivism. Indeed, it is much the same point that I made in discussing Davidson, approached from another direction. We have already considered why I cannot be mistaken most of the time about my own beliefs and desires. If I could be so mistaken, then I could not learn to use a significant portion of my language. Why should this be so? Taylor's remark about the relative independence of other people and objects helps to make some sense of this. To the extent that I can be wrong about my beliefs and desires I shall have difficulty seeing them as *mine*, and not as independently existing or occurring things like spatio-temporal objects and events. If I were usually mistaken about my hopes and wishes, they would seem, not so much to be my hopes and wishes, but things that happen to my body--similar to muscular spasms or chronic snoring. They would not seem part of me. For knowledge to be knowledge of myself, it must let me distinguish myself from others and from other things. This requires that that of which it is knowledge not be independent to the degree that other people and objects are. It does not require--indeed, it requires that it *not be the case*--that I cannot be mistaken about what first-order attitudes I have. It would not be knowledge, if I could not make mistakes.

Now, it would be hasty to suggest that Davidson holds these views, too; they are an extension of his position. But this extension suggests a further dimension to the variable asymmetry between self-knowledge and knowledge of others. I have considered cases in which first-person knowledge-claims have a certain authority that second- and third-person claims do not. We have also seen that there are cases in which this asymmetry of authority is eroded, because non-interpretive knowledge of others is possible, and cases in which interpretation is needed for the *acquisition* of self-knowledge. But Taylor's way of putting things suggests that another sort of asymmetry can separate self-knowledge from other-knowledge, for self-knowledge, he argues, is linked importantly with human agency.

Taylor's point is that *how* we articulate our attitudes can be as important as the fact that we articulate them at all. Certain ways of describing myself can enable me to act in ways that seemed beyond me when I articulated my wants differently. Taylor gives the case of an obese man who wants to control his desire for rich desserts. There is, says Taylor, more than one way in which this man can describe what he wants. He might see controllable obesity in highly charged moral terms, so that what he wants is the attainment of a kind of virtue:

> As I struggle with this issue ... I can be looking at the alternatives in a language of qualitative contrast. I can be reflecting that someone who has so little control over his appetites that he would let his health go to pot over cream-cake is not an admirable person. I yearn to be free of this addiction, to be the kind of person whose mere bodily appetites respond to his higher aspirations, and don't carry on remorselessly and irresistibly dragging him to incapacity and degradation.[17]

Depending on the sort of person I am, this account of my "yearning" may be what I need to overcome my temptation. But it might also stand in my way, making the problem seem too serious to deal with. I may doubt that I have the moral fortitude to face down this challenge, and I may give into my desire despairingly. In such a case, redescribing what I want can alter my prospects of gaining it.

> I might be induced to see it as a question of quantity of satisfaction. Eating too much cake increases the cholesterol in my blood, makes me fat, ruins my health, prevents me from enjoying all sorts of other desired consummations; so it isn't worth it. Here I have stepped away from the contrastive language of strong evaluation.[18]

Now that the moral weight has been lifted from my shoulders--now that my human worth and dignity are not on the line--I can approach my problem with less self-fulfilling fear of failure. A new articulation for my desire restructures the field of possibilities that lies before me, and that restructuring is also a *change* in my desire and in myself. I have acquired "quite a different kind of motivation."[19]

In such cases my knowledge of myself can open up possibilities for me in a way that my knowledge of others does not--either for myself or others.[20] This asymmetry is also related to another asymmetry which, Akeel Bilgrami has argued, characterizes self-knowledge. Knowledge of myself is presupposed in holding me responsible for my behaviour. My knowing what I am doing, what I believe, what I desire, and so on constitutes a basis for praising or blaming me, whereas my knowing someone else's mind does not normally justify ascriptions to me of

---

[17] *Ibid.*, p.21f.

[18] *Ibid.*, p.22.

[19] *Ibid.*, p.36.

[20] At the very least the self-descriptions that I accept can have ramifications for my agency. Insofar as articulating my attitudes sometimes helps to decide what intentional phenomena I have, my articulation gives me self-knowledge, not just a useful self-description. But not all cases are like this. I shall comment further on whether self-descriptions need be correct to be liberating in §VI below.

responsibility for her behaviour.[21] The connection should not surprise us, since Taylor takes his view to offer "another purchase for the concept of responsibility,"[22] but for Taylor what is interesting about the relation between my articulations and my responsibility is that I can be held responsible for the particular articulations that I choose. How I understand myself reflects upon my character, and I am held responsible for my self-understanding, he says, "within the limits of my capacity to change myself by fresh insight ..."[23]

But once again the asymmetry is a matter of degree, not something absolute. In knowing my beliefs and desires under one description, I can be able to envision and assume available courses of action and commitment different from those I imagined under another description. But my interpretation of (or my assumptions about) the behaviour of another can similarly affect my conception of what is possible for me. Whether I see in another the possibility of a developing friendship can be affected by whether I take her biting remarks to be intentional insults or good-natured ribbing. What possibilities present themselves for dealing with a political opponent will be directly affected by whether I see her as mean-spirited and cynical or well-intentioned but mistaken. And not only *my* prospects are affected by my interpretations of others. If a group shares the interpretation that one of its number is self-centred and untrustworthy, then that person's chances for interaction with members of the group are thereby altered. And the effect can be accentuated, if she learns of this shared interpretation of her attitudes and behaviour. Furthermore, to the extent that I am held responsible for my self-interpretations--a point that Taylor thinks further elucidates the notion of responsibility--there is an additional sort of continuity, for we are also held responsible for what we say about others. None of this is very contentious, but it suggests a continuity with cases of self-knowledge and hints that there is no yawning conceptual gulf between justifying first-person claims and justifying second- and third-person claims.

Bilgrami's proposed link between self-knowledge and responsibility is not completely asymmetrical in comparison with knowledge of others, either.[24] In holding someone responsible for her behaviour we usually presuppose that she knows her own mind and what she is doing. But we also hold people responsible for what others do sometimes. A parent is sometimes expected to know what her child is doing and will in those cases be held responsible for the child's behaviour. Where the other person is "morally dependent," or where the "other" is not a person at all, but, *e.g.*, a pet, the one on whom this dependency rests is expected, to a degree, to know what the other is doing. Indeed, this expectation can extend beyond moral dependants to independent moral agents. We are sometimes viewed as responsible for the behaviour of

---

[21] See Bilgrami (1992). Bilgrami takes this asymmetry to be *the* asymmetry which distinguishes self-knowledge from knowledge of others and would not sympathize with my defence of Davidson.

[22] Taylor (1985), p.35.

[23] *Ibid.*, p.39.

[24] A similar point was made to me by Tilman Lichter. See Bilgrami (1992).

family-members, mates, friends and political allies. And I can similarly feel proud or ashamed, emboldened or embarrassed by the beliefs, sentiments and behaviour of one close to me.

I am not denying that there is an asymmetry. I am simply claiming that the asymmetry afflicts certain cases and not others, and that there is no underlying essence of self-knowledge that sets it apart from knowledge of others, or knowledge of the world, for that matter. My complaint is with the procedure that insists that, because certain kinds of cases are important and perhaps more common than other kinds of cases, those other kinds of cases are not as important.[25]

Taylor's account of the relevance of self-descriptions for agency suggests a political dimension to my articulations as well, for I can speak only the language I have learned. To the extent that I have at my disposal a limited vocabulary with which to formulate my beliefs and desires, there are some formulations that elude me and which, had I access to them, might help restructure my sense of what is possible in a radical way. Conversely, if it is in the interests of others that I not discover these new formulations, it will also be in their interests to urge on me certain self-descriptions, or paradigms of self-description, to the exclusion of those liberatory ones that I might otherwise embrace. So, we get some sense from this picture of how ideology can place constraints on our agency. But how absolute are these constraints? And what does the possibility of this sort and other sorts of self-ignorance tell us about the nature of the self?

III. Self-Ignorance and Self-Division

My discussion in earlier chapters has made implicit use of a notion of *self-unity*. The philosophical views that I have criticized suffer, I have said, from epistemic *neuroses*.[26] They entertain theses--especially sceptical ones--that they cannot afford to take seriously, if their own positive programmes are to seem compelling. My critical use of 'neurosis' invokes a normative valourization of a kind of unity of the self. In its use as critical metaphor 'epistemic neurosis' is a diagnosis of philosophical views that show a subtle inconsistency in their concomitant needs to affirm and deny sceptical worries. I also used the norm of self-unity in Chapters 3 and 4, claiming in criticism

---

[25] I may seem vulnerable here, because I have neglected non-intentional states such as pain in my analysis. Let me say in passing, but without substantial defence, that I disagree with Wittgenstein's claim that "It can't be said of me at all (except perhaps as a joke) that I *know* I am in pain" (PI §245). Wittgenstein's point, I take it, is that being in pain is something about which I cannot be mistaken and about which I can have no doubt. Since knowledge presupposes the possibility of mistakes and doubt, I cannot be said to know that I am in pain. "What is it supposed to mean--except perhaps that I *am* in pain" (PI §245)? My inclination is to respond by saying that I *can* make mistakes about being in pain--though I seldom do--and that if I cannot easily doubt that I am in pain, then this is because my knowledge of prolonged pain is extremely reliable (for reasons connected with evolutionary success) and my less reliable knowledge of brief pain seldom issues forth in doubt, because there isn't enough time.

[26] For more on this see §VII below.

of relativism that a "bicultural" person would face a kind of internal logical incommensurability, if the relativist's view were granted. This example hints at a link between self-knowledge and self-unity, a link that is visible in my picture of first-person authority and the justifiability of trust-relationships. As well, making sense of forms of self-ignorance like self-deception, as Davidson says, requires "the idea that there can be boundaries between parts of the mind ..."[27]

The idea of self-division, however, has received more attention from "continental" thinkers, especially those interested in radical theory, and their Anglophone admirers and critics. And just as the dichotomy of objectivity and subjectivity has marked ideas of self-knowledge, so it has affected ideas of the self, as the combinations of subjectivity and objectivity in the Cartesian dream of self as substance and the Humean account of self as mere appearance might lead us to expect.

In Chapter 5 I proposed some likely sources of the temptation that can lead radical theorists to conflate realism with scientism. One source lay in strains of radical theory which, in explaining the power of oppression, mystify the prospects of resistance. This tendency, I suggested, was evident in a sceptical objectification of the self in the work of Marcuse and Althusser. For both, but for distinct reasons, the possibility of self-knowledge is curtailed by ideology. Industrial society, says Marcuse, has developed subtle forms of social control that tend to obviate the need for violent control by leading the oppressed to find fulfilment in the conditions of their oppression. Control requires inculcating a self-ignorance in which the oppressed acquire false desires and delusions about their real wants and needs.

In Althusser's case a similar self-ignorance is achieved, but as a general effect of ideology, whose role is to "'constitut[e]' concrete individuals as subjects" (LP 171). Although Althusser raises the hope of "bad subjects" (LP 181), who can resist the control of ideology, it is unclear that resistance is possible in his picture. His subject of ideology is constructed to be oppressed. As Benton says, "there is no basis for 'interpellations' of oppositional forms of subjectivity."[28]

We saw in Chapter 5, §IV that one reaction to this threat to self-knowledge--a threat echoed in Burge's anti-individualism--is to stress the subjective pole of our now-familiar dichotomy. To avoid portraying the oppressed as benighted pawns we must accept a strong form of first-person authority. And since subjectivism is "certainly not" (RVR 170) coherent, we must settle for first-person-*plural* authority--the relativistic 'we' of the "community of resistance" (RVR 179).

The problems with this view are considerable, as we have seen. The radical theorist tempted by relativism tends to accept the dichotomy of objectivity and subjectivity, seeing metaphysical realism and relativism as the only possible treatments of objectivity.[29] She also tends to conflate realism with *scientism* and responds to the latter by embracing relativism. But the bright star of epistemic relativism does not have enough philosophical fuel to avoid collapsing into the black hole of subjectivism. Only if the relativist can show some crucial difference between the individual and her community, can she rightly

---

[27] Davidson (1985), p.147.

[28] Benton (1984), p.107.

[29] See Chapter 4, §II.

claim that the incoherence of subjectivism does not befall epistemic relativism, and this means that her position ultimately depends on conceptual relativism. On that view, cultures could be, in principle, unintelligible to each other. This amounts, as I argued in Chapter 4, §III, to seeing different cultures as *essentially* different. Thinking that cultures have distinct essences, the relativist drives a wedge between "social and individual accomplishments," between "collective" and "individual, representation[s]" (RRSK 32). But I argued in Chapter 3, §§II-III that any reason for holding a culture or conceptual scheme to be in principle incomprehensible to us would also be a reason for denying that its members were language-users. All this, I claimed,[30] is compatible with admitting that cultures can have different sets of concepts that are incommensurable in *practice*, if their motivating interests are incommensurable. Conceptual "schemes"--in a sense less formal than Davidson's--are rooted in different ways of life. So, the epistemic relativist lured by the "ethical-political argument"[31] has no recourse in conceptual relativism, and her response to a muddled conception of "realism" falls prey to the troubles of subjectivism.

But there is another sort of response to radical theory's sceptical objectification of the self, hinted at in some of Althusser's own work. This response seeks escape from ideological bondage by declaring that the self is not "unified" or "centred," but "fractured" or "de-centred," so that part of it might lie outside the influence of the ruling ideology. Let us approach this position by first back-tracking over the terrain of the "unified" Cartesian self.

I have been construing the "Cartesian conception" of first-person authority in the idiom of analytical philosophy: the Cartesian sees self-knowledge claims as incorrigible and infallible. As we saw in Chapter 6, it is not free of controversy to attribute this view to Descartes. He presents his position in terms of *ideas*; judgments, by contrast, he views as unreliable--at least until justification has been found for them in the simple ideas disclosed by analysis. Even then I must be careful not to judge, if my perceptions are not clear and distinct. But Descartes has an analogue for incorrigibility in the simplicity and consequent indubitability and certainty of the *cogito* (and, I argued in Chapter 6, of the current occurrence and simple contents of my ideas). On Gueroult's reading, this immediate certainty is an "exact coincidence between my thought and existence" in which "the object posited is nothing but the subject."[32] In the *cogito* pure thought is immediately and fully *present* to itself. Nothing intervenes and nothing is hidden. Pure thought in the *cogito* is "my pure self, which has nothing in common with my individual, personal, concrete self, which can only be captured empirically ..."[33] But even if this pure self is--again in Gueroult's (translated) words--a "self common

---

[30] Chapter 3, §§III & V.

[31] Chapter 5.

[32] Gueroult (1984), p.27.

[33] *Ibid.*, p.32.

to all men,"[34] it remains the epistemic and metaphysical focus on which the empirical self is centred and in which it has its unity.

I think that it is a similar reading of the *cogito*, entailing that the "centred" or "unified" subject can be fully present to itself as pure thought, to which Althusser is reacting. Indeed, the "subject" of Althusserian ideology is like the Cartesian subject, with important qualifications. First, Althusser's Marxism is a materialism; so, his "subjects" have, not Cartesian "ideas," but "*material actions inserted into material practices governed by material rituals which are themselves defined by the material ideological apparatus ...*" (LP 169). Secondly, this centred subject is not taken to be metaphysically centred in the way that the Cartesian empirical subject is.[35] Rather, it is "constituted" by ideology, by the "imaginary distortion" of the individual's relation to her real conditions of existence.

But the most important contrast between the ideological subject and the Cartesian subject stems from their similarity. For Althusser, part of the function of ideology, is to render, *as though indubitable and necessary*, contingent features of the real conditions of existence that obtain at a certain point in history. It does so through subjects who--like the Cartesian subject in the *cogito*--find their identity in those conditions so represented. Doubting the necessity of the current socio-economic order thus seems like doubting the very conditions of one's own existence. Ideology, says Althusser, "imposes ... obviousnesses as obviousnesses which we cannot *fail to recognize* ..." (LP 172). But, unlike the *cogito*, this intuitive awareness of the "self-evident," "simple" "truths" of ideology is an illusory experience.

Now, I have raised criticisms of Althusser's view. But just as he thinks that ideology needs the category of the centred subject--"the category of the subject is constitutive of all ideology" (LP 171)--he also holds in "Freud and Lacan" that the "real subject" is

> ... de-centred, constituted by a structure which has no 'centre' either, except in the imaginary misrecognition of the 'ego', i.e. in the ideological formations in which it 'recognizes' itself. (LP 218f)

What is it for the self to be "de-centred?" The de-centred self, as I have taken the metaphor, is not at its core a thinking substance, pure thought. But also, it cannot be completely and immediately present to itself as pure thought. That is--judgments about self-knowledge are not incorrigible; I can be mistaken about myself in any particular case. On this model disunity and self-ignorance go hand in hand.

However, these remarks do not distinguish Althusser's views from my own, since in denying the incorrigibility of first-person beliefs, I allow the possibility of forms of self-ignorance. Where does the difference enter? It makes its appearance *via* the *sort* of critique of Cartesian consciousness that Althusser accepts, an account quite different in the extent of its claims from the one I have endorsed:

---

[34] *Ibid.*

[35] If Althusser cannot account for "bad subjects," the subject effectively *is* centred, much as the conceptual schemes of conceptual relativism rest on the metaphysical realist's account of objectivity.

> ... Freud has discovered for us that the real subject, the individual in his unique essence, has not the form of an ego, centred on the 'ego', on 'consciousness' or on 'existence'... (LP 218)

Decisive for Althusser in deciding whether the self is "centred" is the existence of the Freudian unconscious--according to his reading of Lacan's account of the Freudian unconscious.[36] The unconscious, for Althusser, seems linked with the ideological subject's obliviousness to her constitution *as* subject *by* ideology. We ideological subjects find our identities in the "obviousnesses" of ideological depictions of material conditions. A threat to current social relations seems a threat to us, and by the same bond we come to feel that society "could not get on without us,"[37] that we are "concrete, individual, distinguishable and (naturally) irreplaceable subjects" (LP 173). But--if talk of degree is permissible--to the extent that I find my identity in ideology I cannot see that I am constituted as subject by ideology. Only if I remain unconscious of it, can it constitute me as a subject. As Eagleton explains, self-ignorance is not only possible on this view, but necessary for my constitution as subject:

> [T]he concept of the unconscious means that the forces which determine our being cannot by definition figure within our consciousness. We become conscious agents only by virtue of a certain determinate lack, repression or omission, which no amount of critical self-reflection could repair. The paradox of the human animal is that it comes into being as a subject only on the basis of a shattering repression of the forces which went into its making.[38]

Now, I have said that this rejection of the centred subject points to a possible egress from the prison of ideology in which Althusser's Marxism leaves us trapped. My idea is this: if conscious action is my activity as a subject of ideology, and if I am unconscious of my determination by ideology, then liberation might depend in one of two ways on the unconscious. It might require bringing what is unconscious to the surface by means analogous to therapy. Or, it might require "unconscious activity"--an unhappy term, which I shall cheer below.

The first option has been thought to have something to recommend it. A decade before Althusser Marcuse lamented that in advanced industrial society "private disorder reflects more directly than before the disorder of the whole."[39] His influence is present in Habermas' claim that *Ideologiekritik* and psychoanalysis seek to resist "systematically

---

[36] Eagleton says that Althusser misreads Lacan. See Eagleton (1991), p.144f.

[37] *Ibid.*, p.143.

[38] Eagleton (1991), p.141.

[39] Marcuse (1962), p.xvii.

distorted communication."[40] Althusser, too, regards Marxist science as capable of critique--if it becomes a "subject-less discourse" (LP 171)--but as we have also seen, this possibility seems unintelligible, given the pervasive influence of ideology in the Althusserian scheme.

However, the phrase "subject-less discourse" might suggest a way of understanding the second option mentioned above. If conscious or reflective activity is possible only in and through ideology--if the recognition of oneself as subject is possible only in and by the imaginary distortion of one's relations to the material conditions of existence--then perhaps the key to liberation from ideology lies in an activity that *embraces* the unconscious. What could this mean?

Here is a suggestion. One facet of Lacan's work that Althusser admires is its linking of the unconscious to language.[41]

> ... Freud studied the 'mechanisms' and 'laws' of dreams, reducing their variants to two: *displacement* and *condensation*. Lacan recognized these as two essential figures of speech, called in linguistics metonymy and metaphor. (LP 207)

Lacan's contribution, thinks Althusser, was to show that linguistic categories apply to the unconscious, so that there can be a "science" of the unconscious, modelled on structural linguistics. This "science" might examine how the unconscious emerges in language in the form of metaphors, puns, unintended meanings, slips of the tongue, etc. My words, on this account, have a multiplicity of meanings that I do not consciously intend. Derrida pursues the idea, writing of a "structural unconsciousness" of language, which hinders "conscious intention" from exhausting the meanings of utterances. To determine meaning fully, he says, a conscious intention "would at the very least have to be totally present and immediately transparent to itself and others".[42]

Such considerations might tempt us to see in writing a kind of unconscious radical discourse, exploiting multiple-meanings to subvert the ideological dominance of customary reflective discourse, which pretends that its meanings are clear and unequivocal. One can detect a

---

[40] The term is from Habermas (1970b). See Habermas (1971), pp.214-300 for a lengthy discussion of the comparison.

[41] It's not clear to me how Althusser's essay on Freud and Lacan stands in relation to his essay on ideological state apparatuses. He assimilates ideology to Lacan's category of the imaginary, and on his reading Lacan opposes to the imaginary order the Symbolic Order of language. With the introduction of language the Lacanian male child's illusion of self-immediacy is disrupted (apparently in the form of the father's prohibition on incest), and its desire to be both subject and object together (*i.e.*, united with its mother, from whom it had previously been unable to distinguish itself) is driven into the unconscious. If we carry out the analogy, Marxist science seems to be the logical analogue of the symbolic order of language, and that would suggest that science forces into the unconscious a desire of the ideological subject to be united with social structures. In that case therapy might function to help would-be revolutionaries cope with their unconscious bourgeois desire for immediate self-presence.

[42] Derrida (1988), p.18.

gesture of this sort in some of Derrida's writings and in the work of some French feminists. "Write your self. Your body must be heard. Only then will the immense resources of the unconscious spring forth,"[43] says Cixous. And Althusser's remarks about the "the ambiguity of the term *subject*" might be taken in this spirit: "*There are no [Cartesian] subjects,*" he writes, "*except by and for their subjection*" (LP 182).

The problem with this path is that requires a *reflective* awakening of the hidden senses of utterances.[44] Derrida goes to great lengths to bring out the attitude that he takes to be latent in Austin's focus on "serious" uses of performatives to the exclusion of utterances used "not seriously,"[45] *e.g.*, on stage or in poetry. And this bringing to awareness seems to have as much to do with *Ideologiekritik* and therapy as with giving free rein to the unconscious of language. Moreover, the idea that a *non*-reflective discourse of the unconscious might have play a useful role is hard to distinguish from self-ignorant "business as usual." If the point of such a discourse were not to bring the sublimated into consciousness, but simply to give the unconscious free rein, how would we know when that had been successful?

For the "de-centring" of the subject to avoid these difficulties, it must allow (1) some kind of reflection on the "unconscious" (or a like category)--*i.e.*, an analogue for "therapy" and (2) the possibility of getting "outside" ideology, or at least "outside" the ruling ideology without losing the possibility of reflective, critical practice. Althusser's remarks that we can observe the imaginary distortions of ideology "if we do not live in its truth" (LP 164), that there can be "bad subjects" (LP 181), and that the "structure" which constitutes the subject "has no 'centre'" (LP 219) itself all suggest that he has these points in mind, but that they can be squared with his approval of the "subject-less discourse" of Marxist science is questionable. Althusser needs to bring the unconscious to consciousness in order to escape ideology, but he needs critique that has already escaped ideology, if he is to bring the unconscious to consciousness.

IV. Multiple-Selves

The Althusserian de-centred subject lacks the unity ascribed to the Cartesian subject, centred on the identity of subject and object in the *cogito*. It also seems to lack any clear chance of liberation or self-knowledge. This has led to criticism of the views of Althusser and others whose work has been influenced by French structuralism, but it has not prevented the idea of a "de-centred" or "fractured" subject from exercising a notable influence in some circles of Anglophone literary theory and radical philosophy (to name but two fields).

A likely reason for this influence in the English-speaking academic world lies in the practical importance for radical thinkers of what has become known as the "politics of difference." Recent decades have

---

[43] Cixous (1976), p.880.

[44] I am not saying that this is necessarily a problem for those authors I have mentioned--only a problem insofar as one construes them, in the limited way I have suggested, as engaging in a kind of unconscious discourse which remains unconscious.

[45] Austin (1962), p.22. See Derrida (1988), pp.13-19.

seen an increasing sensitivity to many cultural and other differences within North American and other societies, and feminists have found efforts to build a united, international women's movement complicated by the diversity among women. If one sees the self as conditioned or constructed by social factors, it is a small step to the idea that the self is as multiple or plural as the society in which it is situated.

Whether multiplicity should be treated in terms of a quasi-Freudian unconscious is another matter. Some thinkers have adopted a view that sees divisions of the self as more like the dissociative states of Multiple Personality Disorder than a Freudian or Lacanian unconscious. Thus, Maria Lugones argues for an "ontological pluralism" (SAAO 502) according to which "the self is not unified but plural" (SAAO 503). "In giving up the unified self," she writes,

> I am guided by the experience of bicultural people who are also victims of ethnocentric racism in a society that has one of those cultures as subordinate and the other as dominant. These cases provide me with examples of people who are very familiar with experiencing themselves as more than one: having desires, character, and personality traits that are different in one reality than in the other, and acting, enacting, animating their bodies, having thoughts, feeling the emotions, etc., in ways that are different in one reality than in the other. (SAAO 503)

This summary is reminiscent of the American Psychiatric Association's current diagnostic criteria for Multiple Personality Disorder (MPD):

> A. The existence of two or more distinct personalities or personality states (each with its own relatively enduring pattern of perceiving, relating to and thinking about the environment and one's self).
> B. Each of these personality states at some time, and recurrently, takes full control of the individual's behaviour.[46]

Important to Lugones' story, but absent from the American Psychiatric Association's criteria, is that the divisions of the self in question arise, not just from cultural difference (no part of the definition of MPD either, of course), but also from oppression.[47]

The suggested link between oppression and self-division is not entirely new. Marx's *Economic and Philosophic Manuscripts of 1844* use Hegel's idea of alienation, filtered through Feuerbach's writings, to

---

[46] Quoted in Giancarlo (1991), p.95. We might charitably see the lack of parallel structure here as revealing a plurality of the self.

[47] This does not mean that particular individuals do not play roles in enforcing political oppression. Nor does it mean that multiple-personality disorder has no political significance--the power of adults over children is a political issue, and the fact that the overwhelming majority of MPD sufferers are female, perhaps because girls are sexually abused in greater numbers than boys, raises question about the effects of male dominance on women's mental health. Ralph B. Allison suggests--though this estimate is no longer current--that about 85% of reported MPD cases are women. See Allison (1977), p.vi.

describe the worker's relation to the process and result of commodity-production. Under capitalism, says Marx, the worker is alienated from the "human essence," which lies in the conscious, free, social, productive activity that capitalism obstructs on his view (MEW S.1 510-22).[48] And while talk of a human essence is absent from his work after 1845,[49] *Capital* still analyzes the "fetishism of commodities," in which the worker's value and social ties are displaced onto the object of commodity-production and exchange (MEW XXIII 85-98).[50]

Cultural subjection does not play a large role in Marx's analyses--either in 1844 or 1865. But interest in the divided experience of bicultural people has a history, too. Writing in *The Souls of Black Folk* in 1903, W. E. B. DuBois described his experience of the "double-consciousness" of being black in America:

> The Negro is a sort of seventh son, born with a veil, and gifted with second-sight in this American world--a world which yields him no true self-consciousness, but only lets him see himself through the revelation of the other world. It is a peculiar sensation, this double-consciousness, this sense of always looking at one's self through the eyes of others, of measuring one's soul by the tape of a world that looks on in amused contempt and pity. One ever feels his twoness--an American, a Negro; two souls, two thoughts, two unreconciled strivings; two warring ideals in one dark body, whose dogged strength alone keeps it from being torn asunder.[51]

This lack of "true self-consciousness," which is also a "double-consciousness," is a complex phenomenon. It seems that DuBois equates the "twoness" with a kind of self-ignorance. But it is not simply a lack of *all* and *any* self-consciousness, for the African-American can still "see himself through the revelation of the other world," the world of the powerful. Nor is it a total "constitution" of the self by "ideology," but a "twoness" of "souls" and "unreconciled strivings," threatening, but resistible. Yet, for DuBois some kind of self-unity, "true self-consciousness," is a desirable and (one supposes) possible goal, though the unity seems not to be that of the "centred" subject:

> The history of the American Negro is the history of this strife--this longing to attain self-conscious manhood, to

---

[48] See Marx (1977), pp.55-67.

[49] --Which is not to say that he has completely escaped from the idea of a human essence. This point was, I think, a tacit consequence of Teresa Brennan's recent lecture on attempting to rework the labour theory of value (Brennan, 1993).

[50] See Marx (1967), I, pp.71-83.

[51] Quoted in West (1989), p.142.

> merge his double self into a better and truer self. In this merging he wishes neither of the older selves to be lost.[52]

To Lugones, in apparent contrast to DuBois, self-unity does not seem desirable--be it possible or not. She cites, among other sources, the work of the lesbian Chicana writer, Gloria Anzaldúa. Reflecting on her experience in the political and cultural "borderlands" of Texas and Mexico, Anzaldúa celebrates the consciousness of the "new *mestiza*":

> The new *mestiza* copes by developing a tolerance for contradictions, a tolerance for ambiguity. She learns to be an Indian in Mexican culture, to be Mexican from an Anglo point of view. She learns to juggle cultures. She has a plural personality, she operates in a pluralistic mode--nothing is thrust out, the good the bad and the ugly, nothing rejected, nothing abandoned. Not only does she sustain contradictions, she turns the ambivalence into something else.[53]

This "tolerance for ambiguity" and for "contradictions" is the means by which the new *mestiza* "copes" with her experience of oppression. For Anzaldúa, "Rigidity means death."[54] To deny one's plurality is to open oneself to "the enemy within."[55]

Lugones finds such remarks suggestive of her own rejection of the "unified self" (SAAO 503). According to this "ontological pluralism,"

---

[52] Quoted in West (1989), p.143. As Anthony Appiah has argued, DuBois seems unable to tear himself away from a kind of racial essentialism; so, to the extent that the self is conditioned by its "race" (a category which, Appiah argues convincingly, is illusory), it is not obvious what sort of "merging" DuBois has in mind. I do not want to pursue this issue here; my aim is simply to draw attention to part of the lineage of the idea of self-division in the context of critique. See Appiah (1985). See also West (1989), pp.138-150.

Some remarks from DuBois's third autobiography are interesting in the present context. "Who and what is this I," he asks, "which in the last year looked on a torn world and tried to judge it?" Reflecting on his earlier autobiographies, written when he was 50 and 70, he says:

> One must ... see these varying views as contradictions to truth, and not as final and complete authority. This book then is the Soliloquy of an old man on what he dreams his life has been as he sees it slowly drifting away; and what he would like others to believe.

But despite acknowledging his fallibility and perspective he believes that he will be "near enough" the end of a full century before he dies "to speak with a certain sense of unity." See DuBois (1968), p.12f.

[53] Anzaldúa (1987), p.79. For related themes with an emphasis on the importance of geographic location to personal identity see Martin and Mohanty (1986).

[54] Anzaldúa (1987), p.79.

[55] *Ibid.*

divisions of the self should be *preserved*, and hopes for liberation from ideology and oppression depend on this very preservation. One's cultural plurality, thinks Lugones, is crucial to overcoming one's subordination. There is a tension in radical theory between depicting the ubiquity of oppression and retaining the possibility of overcoming that oppression, and, she maintains, only an account of the self that does not treat it as essentially unified can cope with this tension. Without internal diversity oppression must seem total and inescapable, because it is "unclear how the self can be unified and contain and express both a liberatory and an oppressed consciousness" (SAAO 501). What she calls "the ontological possibility of liberation," is said to "depend on embracing ontological pluralism" (SAAO 502).

Phrases like "ontological pluralism" and "ontological possibility of liberation" bring to mind a problem that I raised in Chapter 3. If the thesis of radical incommensurability--*i.e.*, that it is in principle impossible for members of the one culture to fathom the members of the other culture and *vice versa*--were true, then the bicultural person would face an insurmountable logical barrier in trying to understand her A-cultural self from the standpoint of her B-cultural self. She would, it seems, be *two metaphysically distinct selves*, each unable to communicate with the other. She would have two "souls" or "strivings," not just "unreconciled," but *irreconcilable*.

An important aim of Chapter 3 was to show that sense can be made of the idea of conceptual incommensurability, but only if conceptual differences are rooted in practical differences. Now, if a kind of practical incommensurability is possible for conceptual schemes and cultures, then the case of the person who lives across cultures or sub-cultures in the cultural "borderlands" raises the real possibility of a practical, internal incommensurability of the self.

But what barriers to the commensuration of a culturally divided self does Lugones mean to designate? Logical barriers, or pragmatic ones? Lugones' talk of "ontological" pluralism suggests the former. But then her position threatens to collapse into an internalized form of the self-defeating logical incommensurability of conceptual schemes. Rorty understands Lugones this way, criticizing her metaphysical language as a sign that she rejects the "desirability of harmonizing one's various roles, self-images, etc., in a single unifying story about oneself." To Rorty, by contrast, "Such unification ... seems ... desirable."[56] But before rushing to agree with Rorty's interpretation, I think we should explore the disunity of the self Lugones has in mind and ask whether any conflict arises between her position and the normative ideal of unity invoked in my denial of logical incommensurability.

Some of the vocabulary that Lugones explicitly *rejects* seems cast off for reasons more in tune with recognizing practical divisions of the self, not metaphysical ones. There is no "transcendental self," she says, that wears its personae as "masks" and "is distinguishable

---

[56] Rorty (1991b), p.12n22. Rorty's paper does a good job of exploring some of the affinities between his position (which is not so far from the one I have been taking) and strains of contemporary feminist theory. However, he seems to think that his position rules out any role for ideological critique (a point on which I must differ from him--see Chapter 5, §IV). See *ibid.*, *passim* and Rorty (1989), p.59n15. For criticism of this and related features of Rorty's views see Fraser (1989), pp.93-110 and Wilson (1992).

from them" (SAAO 506). So, it also is reasonable to expect her to deny that there are *two* or *more* "transcendental" selves, instead of one.

But other claims that she makes are reminiscent of Althusser's. Persons are "construct[ed] or constitute[d]" by "structures"--

> patterned arrangements of role-sets, status sets and status sequences consciously recognized and regularly operative in a given society and closely bound up with legal and political norms and sanctions. (SAAO 506)

Maybe Lugones is suggesting that, since there is a number of distinct structures ("ideologies" in Althusser's sense), each individual is constituted as a number of distinct "subjects" by those distinct "ideologies."[57] But if the only contrast between Lugones and Althusser is that she treats the individual as a plurality of "subjects," each constituted by a different "structure," then her view seems to throw up logical barriers among an individual's selves.

However, Lugones' position departs from Althusser's in two important ways. First, whereas "ordinary life presents structures as systematic, complete, coherent, closed socio-political-economic organizations or normative systems," she insists that "structures are not closed" (SAAO 505).[58] Here is a hint of the possibility that one's different, constructed selves are not "closed" either, but pragmatically incommensurable. Secondly, Lugones seems to see the possibility of agency under oppression as lying, not simply in the plurality of the self, but in an *awareness* of this plurality. Thus, speaking of the different "worlds" that a bicultural person may inhabit, she says, "It is very important whether one remembers or not being another person in another reality" (SAAO 504). It is to the degree that one can be made to forget one's life in another "reality" that an oppressive "reality" seems inescapable, exhaustive of *reality*. Remembering one's other selves, however, can bring to mind the possibility of an alternative to the current order:

> [T]he liberatory experience lies in this memory, [it lies] on these many people one is who have intentions one understands because one is fluent in several "cultures," "worlds," realities. One understands herself in every world in which one remembers oneself. This is a strong sense of personal identity, politically and morally strong. (SAAO 504f)

The very process of remembering, thinks Lugones, is an escape from the rigidity of "structures" and "into the limen," the "place in between realities" (SAAO 505). Simply to remember one's other selves is to get outside one's constitution as "subject" by an Althusserian "ideology."

The parallel with MPD, suggested above, is quite striking here. The personalities of MPD-sufferers are often reported to have no memories

---

[57] Althusser would resist any such treatment, if it amounted to treating the "unconscious" as a "second consciousness," a move which he classifies amongst "ideological misunderstandings" (LP 208).

[58] As we saw, Althusser does say that structures are themselves "de-centred," but, as we also saw, it is questionable whether some of his other views are compatible with this remark.

or awareness of one another. Joan Frances Casey, in her autobiographical account of life as a "multiple,"[59] describes her "Jo" personality as academically gifted, but incapable of emotional involvement and so convinced of her own physical repulsiveness that she could not look into a mirror without losing consciousness to another personality. Prone to bouts of amnesia that could last for months, Jo was unaware of Casey's other personalities, denying their existence, when Casey first entered into therapy, and then regarding their tales of past abuse at the hands of her parents as lies.[60]

What made Casey's therapy successful was the presence of other personalities who were conscious of each other. Renee, whose viewpoint is that of the narrator in Casey's book, describes to her therapist her ability to know what most of the other personalities were thinking or feeling without herself so thinking or feeling, and without having the talents and abilities of the other personalities:

---

[59] Casey and Wilson (1991).

[60] That MPD is a genuine disorder is less controversial amongst mental-health professionals than it once was. In the past decade the study of "Multiple Personality Disorder" has become a growth-industry for therapists and theorists. The expansion of this field is attested to by the astonishing proliferation of scholarly and clinical research being produced on the topic. In 1981 *Psychological Abstracts* listed only three articles on the phenomenon, in 1984--twenty-one. By 1991 this number had increased to 113, and the American Psychiatric Association recognized MPD as a genuine mental disorder in 1980, updating its diagnostic criteria in 1987. See Giancarlo (1991), p.95. The field has had its own journal (*Dissociation*) since 1988.

Of course, acceptance of the MPD classification is not universal--a fact suggested by the recent exchange on the letters-page of *The Sciences*, between the prominent MPD-researcher, Frank W. Putnam, and his critic, Harold Merskey. See Putnam (1993) and Merskey (1993). But as Hacking argues (Hacking, 1991, p.842f), much debate over whether multiple-personalities exist tends to conflate a number of distinct questions: (a) a question about whether multiple-personality is a "real, objective state," (b) a question about whether there is "a set of core behavioural criteria" which could pick out "one possible kind of behaviour" to be found "at least across a substantial part of Western culture" without any necessary commitment to MPD being a "connected with some ... identifiable neurological, biochemical, or physiological abnormality," and (c) a question about the instrumental effectiveness of treating patients as if multiplicity were "a real part of their character," again without any necessary commitment to positive answers to either of the first two questions.

Hacking regards (a) as currently unsupported, though worthy of investigation, should one answer (b) in the affirmative. And he sees (c) as an issue to be settled by patients and therapists. To (b) he offers a cautious, but affirmative answer, based on an examination of cases of multiples "entirely outside the 'canon' published in the literature of the [MPD] movement" (*ibid.*, p.844). In Hacking's words, MPD is "a way to be crazy, at least in an industrial/romantic, Protestant society" (*ibid.*). For a brief but intriguing critical history of the notion of multiple-personality see Hacking (1986).

> "I go inside, but I'm nearby," I said. "It's like I'm at the back of a theatre, watching a play. But if I get bored watching, without really meaning to, I leave the theatre and go inside another room in my head. Sometimes a question or some sort of warning bell calls me out. When I come back, I know what happened, but I don't feel like I really did anything."[61]

Going "into the limen" and developing a "tolerance for ambiguity" are things that a therapist may well encourage a multiple-personality to do. She may look for a personality amongst the many she meets, who is aware of the other personalities and does not try to deny that awareness, one who can serve as an "inner self-helper"[62] or manager.

But there is a tension emerging here, for simply remembering one's other selves is not regarded as enough, if the MPD-patient is to be treated successfully. Therapist and patient also try, albeit slowly, to knit together some kind of unity from the separate strands present; indeed, the managerial personality herself is a kind of unity inasmuch as--to employ Kantian terminology--she binds together her separate consciousnesses in a single consciousness. By analogy, going into the limen, as Lugones recommends--be it sufficient to recover agency or not--might be seen as a move in the direction of unity, or, as Lugones says, of "personal identity," not a step away from it. Why, then, does she recommend against attempts at promoting self-unity? My account of self-knowledge provides a way of answering this question.

V. Self-Descriptions, Agency and False Unities

In Chapter 6 I ascribed to Kant a version of the claim that self-knowledge involves an articulation of one's intentional phenomena that "does not leave its object unchanged."[63] For Kant, I said, the self is constituted by a general empirical unity of apperception, or better,

---

[61] Casey and Wilson (1991), p.21. Renee reveals that some of the personalities could carry on conversations with her, and after extended therapy bizarre instances of "co-consciousness" became possible:

> As other students recited, or when the professor was lecturing, I scribbled notes on my tablet right-handed, with the casebook opened to my left. Able to use both hands equally well, Kendra often made notations in the casebook with her left hand at the same time that I was taking my notes. (*Ibid.*, p.212.)

Despite such displays of intellectual virtuosity, however, each of these personalities suffered from major deficiencies. Anxious for the approval of others, Renee was skilled at multiplying sexual relationships with men and at forming instant friendships. But she expected her relations with others to be temporary. When faced with disapproval she would retreat to "another room in [her] head," leaving some other personality to cope with criticism or perceived rejection.

[62] Giancarlo (1991), p.98f. See Casey and Wilson (1991), p.57.

[63] Taylor (1985), p.36.

an empirical unity *of* empirical unities of apperception--a binding together of representations, but also of my prior judgments, my previous "bindings" of representations. There is no limit to the recursive application of such "synthesis." As we saw in Section II, such arcane talk of binding and synthesizing can be succeeded by talk of "articulating" my intentional attitudes. The ability to articulate my intentional phenomena constitutes my capacity for self-knowledge, and articulation sometimes "does not leave its object unchanged."

This latter fact about articulation suggested a way in which self-knowledge matters to one's capacity for agency. How I think of myself, what my self-descriptions are, structures the field of possibilities before me.[64] To recall Taylor's example: whether a man, obese from over-eating, finds that he can resist the temptation of a rich dessert may hang on whether he describes the problem in weighty moral terms or in the prudential terms of maximizing expected utility. If he sees his obesity as rooted in a lack of self-control, in a flawed character, then this may so inflate the problem that he despairs at the prospect of having to prove his moral worth. But if he reckons that losing weight will enhance his health and his prospects for the enjoyment of other activities, then the problem may seem just a matter of applying the right techniques, without fear of failure and self-betrayal.

Which strategy, among others, is more effective depends on the sort of person facing the problem. Sheila Mullett offers another dietary example that illustrates this point.[65] Mullett tells of a young woman who found herself suffering from bulimia and, at the level of expected utility, understood the problem of her food-addiction perfectly well:

> [S]he had a solid grasp of all the relevant information and her inferences were sound. She knew that this kind of behaviour was debilitating and she attempted to use her will power to stop it, but she could not.[66]

What proved effective for her was, again, a re-description of herself and her problem, but this time in terms of her moral relationships with others. Mullett quotes from the woman's personal account:

> Bulimia meant for me that my personal life was safe, in a shell, with no other participants ... [it] filled the time that would otherwise have been lonely. I felt connected to no one and it occurred to me that most of my daily routine was revolving around the search for this self-gratification.[67]

---

[64] We should distinguish self-description from self-knowledge, since some self-descriptions are false. This raises the possibility that false self-descriptions might be liberating, and the possibility that self-knowledge might threaten agency. See Section VI for more.

[65] Mullett (1987). Her example is meant to show something broader about the importance of one's personal relations for self-knowledge and agency. She seems to regard Taylor's example as showing that the weighty moral description of my problem will help me to overcome it.

[66] *Ibid.*, p.329f.

[67] Quoted in *ibid.*, p.330f.

She goes on to describe the changes in her life that made a new set of self-descriptions possible. She started to lead a less "self-centred existence," learning to value the talents and successes of others, rather than see them as indicative of her own deficiencies: "I used to practice caring for other people, giving things away until I felt good about it, listening to others ... the list goes on ..."[68]

The central point that Mullett draws from this example is that the changes of self-description that are crucial to agency often depend on prior changes in one's relations with others. How one can interpret oneself can be changed by an alteration in how one stands with others:

> [C]hanges in our experiences are more likely to arise out of changes in our relations with others than out of changes in our vocabulary. The changes in vocabulary *result from* the changed relations.[69]

I am not sure that Mullett tells the whole story here. How one relates to others can be affected by one's self-descriptions, even if these descriptions are affected by one's relations with others. One's self-descriptions and one's personal relations stand with each other in a dialectical relation: each affects and is affected by the other. Maybe some series of events involving others led the young woman to realize that her life centred on "the search for this self-gratification." But this realization--including, we may infer, the belief that self-centredness is not desirable or is excluded by her ideal image of herself, of who she wants to be (or "really" is)[70]--was itself a change that seems to have preceded her recovery:

> All the while I was bulimic my conscience would tell me that something was gravely wrong with me, but I simply didn't know how to change. I figured I could will my way out of bulimia but when I tried it, I never made any positive changes.[71]

It appears that it was, in part, a new articulation of her problem that let her begin "to practice caring for other people." Seeing her behaviour as a sign of weakness gave no counsel but the unsuccessful, individualistic one of trying to be stronger. But seeing her behaviour as self-centred suggested a way out with the help of others. So, while Mullett is right to stress the importance for agency of "changed relations" with others, it remains plausible that changes in self-descriptions are of importance for agency in their own right.

---

[68] Quoted in *ibid.*, p.331.

[69] *Ibid.*, p.337f.

[70] My use of "really" is meant to capture a common way of talking about such problems; it is not an affirmation of the idea of a metaphysically essential self. But, I shall argue later, the fact that there is no *essential* self does not mean that there is no self at all. Self-descriptions can err and can be warranted prior to their articulation (though not always). Self-descriptions can also be instrumentally warranted without being epistemically warranted.

[71] Quoted in Mullett (1987), p.330.

This compromise still allows us to draw the following lesson from Mullett's example. One's self-descriptions and -interpretations do not spring *ex nihilo*. They arise in the context of one's dealings with others in a particular society, at a particular time in history, amid particular personal and cultural influences. Indeed, we saw while examining trust in Chapter 7 that self-knowledge is--in the first instance--modelled on our knowledge of others. One's sense of oneself, of who one is and what is possible for oneself, is affected by one's culture or cultures, by one's station in life, by how one stands with others. So, if one lives in the borderlands, a product of more than one culture, then she has at her disposal, at least potentially, more than one set of vocabulary on which to draw in self-interpretation. But if one's first culture is subordinated to another, one can also come to see oneself through the eyes of that dominant culture.

What does this tell us about Lugones' views? First, if the ways of life of a bicultural person's cultures conflict practically, there can arise cultural incommensurabilities of a similarly practical nature. So, the vocabularies on which she can draw may also conflict. From the vantage of one of her cultures she may have trouble understanding those parts of herself that are drawn to her other culture. This practical barrier may leave her ambivalent about her choices and behaviour, about her desires and beliefs. But secondly, the problem will be intensified, if the two cultures are not merely in conflict, but in a relation of domination and subordination. As the legacies of European encounters with other peoples show, new cultural contact is often far less benign than might be hinted by my account in earlier chapters of comparison, understanding and the gradual erosion of cultural boundaries.[72] Among the most significant of differences within a culture or between cultures are those used to justify the inequitable allotment of power, and these sorts of inequities form the other element of the experience on which Lugones and Anzaldúa draw.

In cases of cultural subordination the potential to draw on one's several cultures in the process of self-interpretation is harder to actualize, because part of the domination of a culture by another is its devaluation. Perceived holders of the "cultural identity" of the subordinate culture may on these grounds be barred from the community of the dominant culture--from the dealings and discourse of power. Often implicit in assigning such a "cultural identity," I think, is a mistaken idea of cultures as bounded by necessary and sufficient conditions. Intra-cultural differences are effectively ignored, and members of the subordinate culture are treated as essentially similar to each other, essentially different from members of the dominant community.[73] Essentializing conceptual schemes, I argued in Chapters 3 and 4, leaves other schemes inscrutable from within one's own, and treating others as incomprehensible in this way amounts to a kind of scepticism about other minds. Similar conclusions apply to assigning a "cultural identity," and it is easy to see how treating cultural others as if unfathomable, even mindless in their behaviour, could mesh with a devaluation and subordination of their culture.

---

[72] See, *e.g.*, Wright (1992). European societies have no monopoly on cultural imperialism, but now their dominance is most noticeable.

[73] I draw here on Chambers (1992), pp.5-24.

Now, consider the effects of such devaluation on the inhabitant of the borderlands. Her dominant culture urges her to see herself--"through the eyes of others," as DuBois says--as essentially different from its "essential" members and, so, not worthy of serious attention. If this control of her self-understanding succeeds, she loses sight of the richness of her subordinate culture (and of herself in that culture), for she sees it as a disqualifying "cultural identity." From the dominant culture's vantage--where she is measured "by the tape of a world that looks on in amused contempt and pity"--she sees herself as less worthy, and because it is a dominant culture, it can command her attention. To get by in the dominant culture she learns to fit the interpretations that that culture would impose on her, to appear only as she is expected to appear, to maintain a certain *invisibility*.[74] But to do this she must hide those "differences" in her background that "make a difference"--to use a popular phrase. Like an MPD-sufferer she forms a personality to fit the interpretations given her by the powerful, and like a multiple-person she can forget her other self, interpreted in the vocabulary of the subordinate culture.[75]

So, a struggle against cultural domination is a struggle over which self-interpretations one can apply. It is a struggle over who one can be and often a struggle to know one*self*, to know what one needs and cares about.[76] Interpreting needs, as Nancy Fraser observes, "is itself a political stake, indeed sometimes *the* political stake."[77]

I suggested earlier that Lugones' "ontological pluralism" might be taken to treat each individual as constituted as several different "subjects" by several distinct Althusserian "ideologies." On Lugones' analysis, the oppressed person ends up in a state like (but less hopeless than) the Althusserian subject's, identifying with the conditions of her subjection. The parallel with Althusser is instructive, since for him the self-unity embodied in the subject of ideology is an illusory unity that is the condition of the individual's subjection. This clarifies why Lugones and Anzaldúa would prefer to inhabit the "borderlands," why they would want to reject a certain sort of self-unity. Their concern, I think, is that insofar as oppression effects divisions of the self, a unity of the self can be a false security--

---

[74] Subverting this "invisibility" through a reinterpretation in the terms of the subordinate culture is a prominent theme of another of Lugones' examples of "ontological pluralism," Ellison (1989).

[75] Casey tells of the emergence of a childhood personality, dedicated solely to entertaining her mother's friends, and later of "Karen," the tidy child who appeared when "her mother said once too often, 'Why can't you be like your cousin Karen?'" (Casey and Wilson, 1991, p.78.) A male personality, "Rusty," first emerged at the age of seven when the author's sexually abusive father said of his penis, "Betcha wish you had one of these" (*ibid.*, p.173). A common topic in the MPD-literature is the *iatrogenesis* of personalities--the creation of personalities who try to please the therapist by feigning recovery.

[76] It need not always be. It may be a struggle to find a better but false self-description, or a struggle to learn what kind of person one wants to be, but is not yet. I'll return to these issues below.

[77] Fraser (1989), p.145.

i.e., it might be only an apparent unity achieved by forgetting one's other "selves," losing one's memory of "being another person in another reality." This sort of self-unity would be a full acceptance by an oppressed person of her oppressors' descriptions. It is the sort of unity that she might attain by forgetting utterly about her life in another reality and even quitting that reality entirely. The coercive nature of her social position would then be effectively hidden and that much more difficult to resist.[78] It would be like the MPD-sufferer allowing one personality to dominate and suppress all the others. Such a multiple-person does not thereby free herself of her old problems. They simply remain unarticulated until the dominant personality can no longer successfully repress them. Similarly, the sorts of unity that Lugones and Anzaldúa reject are specious sorts of unity--"unities" that depend on self-ignorance.

But a justified caution about false unities of the self is no threat to valuing the unity embodied in the possibility of taking one's many selves up into a "single consciousness"--to revert again to a Kantian idiom--or of articulating conflicting beliefs and desires in a way that does not occlude the conflict, but sees the conflicting attitudes as "all mine" and as in need of a resolution that does not consist just in embracing one set and rejecting another. This realization is implicit in Lugones' counsel that the bicultural person go "into the limen," and it hints at a plausible reading of Anzaldúa's claim that "nothing" is abandoned by the new *mestiza*. The new *mestiza*, she says,

> ... can be jarred out of ambivalence by an intense, and often painful, emotional event which inverts or resolves the ambivalence. I'm not sure exactly how. The work takes place underground--subconsciously. It is work that the soul performs. That focal point or fulcrum, that juncture where the mestiza stands, is where phenomena tend to collide. It is where the possibility of uniting all that is separate occurs.[79]

The MPD-analogy might suggest the pragmatic importance of rejecting nothing, of concentrating instead on letting undesirable traits grow into mature, healthier characteristics. Joan Frances Casey's therapist resisted the urge to reject the violently self-destructive "Josie" personality and the misogynistic boy "Rusty"--personalities who, over time and through caring and acceptance, grew out of their undesirable behaviours and views to be integrated into Casey's post-therapeutic personality. Successful integration rested on letting hateful and self-destructive feelings surface and be dealt with directly, through an understanding of their origins in response to childhood abuse--rather than on keeping them repressed, pretending that the anger and fear that they represented were not worthy of acknowledgment.[80]

---

[78] I get the notion of "hidden coercion" from Campbell's account of "hidden compulsion." See Campbell (1979), pp.156-165.

[79] Anzaldúa (1987), p.79.

[80] This is not a universally applied therapeutic technique. In his autobiography (less impressive than Casey's) Henry Hawksworth describes the moment of his alleged reintegration as one in which two of his personalities fight to their mutual death, awakening the young

That a similar policy should be prudent for the culturally divided person is not surprising. When members of one culture dismiss the grievances of members of another culture, the discontent of the latter is seldom alleviated. In addition, the parallel with MPD is very real in the following respect: it is a rare individual who can utterly reject or repress her cultural background. Inasmuch as she is formed under the influence of this culture it will not simply disappear, but its manifestations will linger in her behaviour and her desires, tugging at her as though a compulsion.[81]

VI. Unity: Its Possibility and Its Desirability

Lugones' emphasis on remembering oneself in another "reality" does not amount to embracing the self of the subordinate culture and cutting loose the self of the dominant culture, as if the former were the real, essential self, hidden under the obfuscating cloak of the ruling ideology. One might associate such a move with either a latter-day Cartesianism about the self as *substance* or a kind of relativism which, by flawed reasoning parallel to that of the "ethical-political" argument of Chapter 5, might suppose that the subordinated self could be lifted up, if seen as essentially different from the self designed to satisfy the dominant culture. Once again metaphysical realism and relativism are tacitly in league with each other.

But to say that there is no essential self, or that the self has no essential unity, is not to say that there is neither self nor unity--

---

[80] boy who "fell asleep" as a child and didn't reappear until middle-age. This, at least was the interpretation of Hawksworth's doctor (as Casey's was her therapist's). Perhaps tellingly, Hawksworth notes near the end of his book that, though they were becoming less frequent, he still had MPD-lapses. The question of which self-interpretation will prove therapeutic probably has a strong instrumental component that varies from person to person, but it is tempting to see a longing for Cartesian times in the original, persistent personality of Hawksworth's interpretation. See Hawksworth and Schwarz (1977).

Giancarlo notes that a technique recommended by Frank Putnam is to have the MPD-patient agree with the therapist on a special contract:

> [T]he therapist needs to contract with the client regarding the limits of treatment, the safety of both the therapist and the client, the therapist's privacy, and the safety of the therapist's property. The contract should be negotiated with as many of the alter personalities as possible. (Giancarlo, 1991, p.99)

Putnam cautions, however, that

> ... the alter personalities of MPD patients are not to be mistaken for whole people. Rather they are best conceptualized as discrete states of consciousness with limited functions and dimensions of personality. In aggregate they make up the personality of the patient. (Putnam, 1993, p.4)

[81] I say "as though," lest it seem I am saying that the influence of one's culture is like the alcoholic's need for another drink.

any more than to say that there are no cultural essences is to say that there are no cultures. The point of my discussion of Lugones has been to show that in advocating a preservation of one's many selves and a going-into the "limen" between "structures," Lugones tacitly advocates a kind of self-unity. Let me clarify this remark.

The sort of self-unity that I have looked for in Lugones' account has Kantian roots.[82] Just as the transcendental unity of apperception grounds the fact, for Kant, that I can say of all the representations of which I can become conscious that they are all *mine*, so the memory of one's various selves in different "realities" is linked with one's ability to say that they are all *my* selves. Lest this "transcendental" unity be mistaken for the metaphysical unity that Lugones dismisses as a "transcendental self" (SAAO 506), recall that for Kant the original unity of apperception is a merely formal and general unity on which rests the possibility of the empirical unity of apperception in particular judgments. Transcendental apperception is presupposed by the possibility of knowledge, as the grammatical subject (to use a non-Kantian term) of all my judgments. To mistake this condition for some kind of metaphysical *entity--e.g.*, a Cartesian substance--is to commit the error of the First Paralogism of Pure Reason.

But if no Cartesian substance underlies all my judgments, how can a single "grammatical subject" do so? Each judgment is distinct in time, and the series of my judgments presents only an ever-growing plurality of grammatical subjects. "The 'I' with which we speak stands for our identity as subjects in language, but it is the least stable entity in language ..."[83] Yet what grounds my distinct judgments at times t_1 and t_2 in a way that makes them judgments of one person is the possibility of my correctly judging further that both *are my* judgments.[84]

---

[82] We might also trace her talk of remembering one's other selves to Locke's proposal that personal identity be understood in terms of memory. Locke's mistake, and that of his critics, was to treat memory as a necessary and sufficient condition for personal identity, rather than just as internally related to the concept of a person. See Locke (1964), Book II, Chapter XXVII, §§6-26.

[83] Rose (1982), p.31.

[84] Note that the possibility of such judgments, not any *actual* judgment, constitutes this basic self-unity--though the former will not likely obtain without the latter. Thus, the fact that I have not thought together two of my "consciousnesses" in a "single consciousness" does not mean that I am lacking self-unity as an MPD-sufferer does. I shall not dwell on the sense of "possibility" that is relevant here. Suffice it to say that there is a sense in which the multiple-personality's capacity to unite her judgments and other intentional phenomena into a single judgment is diminished. However, even for the multiple there is more than just a logical possibility of self-unity. Therapy reinforces her capacity for unity. We might also conjecture that it is possible in a weaker sense (still stronger than logical possibility) for two distinct bodies to become conscious of each other's intentional phenomena by way of some futuristic neural patch-cord. But this case falls outside our rough and ready criteria for personal identity, criteria on which I shall comment shortly.

If language is to be intelligible--and I have assumed without apology that it is--then there must be selves, persons--grammatically integrable subjects. The ability to speak a language, as I have insisted, involves the ability to make use of personal pronouns, and there are correct and incorrect ways of using those pronouns. Part of what makes my applications of the pronoun 'I' correct (when they are correct) is that what is predicated of this "I" is done so correctly. How can I and others discern whether this is the case? There is no simple answer to this question. Knowing when such predications are correct is part of one's ability to use personal pronouns. But part of that ability is surely linked with recognizing persons as individual human bodies. "The human body," as Wittgenstein says, "is the best picture of the human soul" (PI II 178). That one person does not feel another's pain or think another's thoughts is rooted in the fact that *this* body is distinct from *that* body. That this human body before me is a person is indicated by (among other things) the fact that from its mouth or its hands or its word-processor issue correct utterances of personal pronouns and other words. Among those correct utterances we will likely find claims of responsibility for past judgments, questions, exclamations, witticisms and so on. Whether these claims are correct is typically shown by whether or not we or others witnessed those judgments--and the like--emanating from that body.

If these locutions seem awkward and alienating, then that is just indicative of the primitive role that the concepts of "person" and "you" and "I" (etc.) play in our lives. We cannot, I think, do without the notion that language has meaning, and we cannot, I think, make sense of meaning without the idea of creatures for whom things can *have meaning*. That, in another formulation, is just the claim I have repeated since Chapter 1: meaning is a normative, intentional notion. And for this reason, more than any other, the possibility of a kind of self-unity has been implicit in my arguments since Chapter 1.[85]

The ideas of personal identity and self-unity are intimately bound up with the ideas of bodily identity and what we might call the "grammatical unity of self-description."[86] But to say so is not to try

---

[85] I do not say that there could be no language without the first-person pronoun. My point is that language needs a "subject"-role, and insofar as language-users bore any significant resemblance to human organisms, the individual body would likely be the standard unit of significance. But there can be corporate subjects to which particular bodies are less relevant--"the team," "the board," nation-states--and we can imagine these playing roles that dominated (superseded?) the roles of "person," "self," etc., given certain ways of life and social organization. Very different organisms--or human organisms with the right technology--capable of collective awareness would probably have a very different system of pronominal reference, but would still have subjects for whom words had meaning, if they remained language-users.

[86] It might be said in Davidsonian terms that the notion of personhood is internally linked to the possibility and necessity of giving holistic, rational interpretations of an organism's behaviour. For a good discussion of this point--including remarks on psychological disorder and "repression"--see Malpas (1992), Chapter 3. As I have noted, I think that Davidson over-emphasizes "interpretation."

to reduce personal identity to bodily identity nor to answer the analytical philosopher's question, "What conditions are necessary and sufficient to distinguish one person from another?" There is likely no answer for that query,[87] and I think that it is a mistake to try giving one. Asking what definitively distinguishes person A from person B is at least as odd as asking what definitively distinguishes two waves in a swimming pool. That the waves are distinct is clear enough when the wave-crests are at opposite ends of the pool, but as they meet, there is simply no fact concerning where one ends and the other begins, or to which wave a given water-molecule belongs.[88] Similarly, the fact that the ship of Theseus can be stripped and re-planked as it crosses the ocean does not show that there is a defect in our concept of "ship" or in the notion of the "same" ship. Nor was the concept of "person" devised to do such "rigorous" work; it is not therefore useless. "The sign-post is in order--if, under normal circumstances, it fulfils its purpose" (PI §87).

One might still want to ask, however, whether anything but the barest version of my analogue for the unity of apperception is either possible or desirable. The perfect unity of subject and object in the *cogito*--on one interpretation--is a unity whose possibility I have already rejected. The moment we get rid of the incorrigibility of self-knowledge claims, we get rid of this ideal unity. But the moment we allow that we cannot be utterly mistaken about our own intentional attitudes, we allow for another sort of unity--the unity given in our particular self-descriptions and self-interpretations.

The question of the desirability of our self-articulations, then, must be understood in one of two ways: (1) Is it desirable that there *be* any self-descriptions? (2) For a particular person in a particular context are the self-descriptions at her disposal desirable? I shall return to the latter later. The former asks, at best, "Should a person try to understand herself?" and at worst, "Should there be such a thing as language, such things as knowledge and justification, such things as cultural practices, or would it be better if there were only 'mere brutes', animals that were not language-users?" I think that we should answer "yes" to both versions of question 1, but it is not a matter that I want to discuss here. The possibility of philosophical activity or of thinking of any sort demands a positive answer to at least the second version, and very likely to the first.

Now, it is important to distinguish two related points that I might be suspected of blurring together here. One is a point about the link between self-descriptions and self-*knowledge*, and the other is a point about the link between self-descriptions and agency. I have tied the capacity for self-knowledge to the capacity for self-description, but I have also insisted that self-descriptions can be mistaken. So, self-knowledge and self-description are not identical--were they, there would be no self-knowledge. But they are internally related. Still, their non-identity raises a question about the relation between self-knowledge and agency. It is tempting to say that agency depends on self-knowledge. Since I cannot be totally in error about my own

---

[87] --As is suggested by the puzzle-cases that philosophers have created for every criterion of identity proposed: same body, same brain, same psychological profile, etc. See, *e.g.*, Parfit (1984).

[88] The example is from Haugeland (1982).

intentional phenomena, and since my agency may be taken to depend on my capacity for self-description, this dependence holds generally. But in particular cases, the only clear dependence relation that my discussion above need be taken to reveal is a dependence of agency on certain kinds of self-*description*. This suggests that some *erroneous* self-descriptions could be liberating or enabling and, conversely, that self-knowledge might sometimes be a threat to agency.

I am obliged by the norm of consistency to admit both possibilities, but doing so does not require that I suppose agency to have no interesting dependence on self-knowledge in many particular cases. This is partly due to the general dependence cited in the preceding paragraph. While I have argued that self-knowledge claims, or first-person beliefs, can be wrong, I remain committed to the idea that I cannot be completely or even in the main mistaken about myself and still be a competent language-user. If I am not a competent language-user, then I have a diminished array of self-descriptions (if any at all) and, so, only a diminished capacity for agency (if any at all).[89]

But I also think that often a false self-description is likely to be liberating, only if it tends over time to become true, or less of a distortion. Let me explain. In trying to cast off habits or urges that we find undesirable, we sometimes identify with certain of our desires in contrast to others. Embracing certain "self-identifying desires"[90] and viewing other desires as "not part of me," I effectively accept a

---

[89] We might see severe and debilitating psychoses as cases in which a person loses her capacity to understand her own words--even while retaining those words. For an example that might be interpreted so see Malpas (1992), p.58f. To illustrate the importance of the principle of charity in interpretation and the thesis of psychological holism, Malpas describes the case of "Smith," who seems to hold wildly inconsistent beliefs and desires about snow, grass, Australia and the South Pole. Where there is so little integration of attitudes, Malpas suggests, we would expect a similar lack of integration in behaviour. Thus, when we decide to hire Smith as a painter, things go awry:

> On being told to paint the garden fence white, she first pulls up the flower beds, heaps flowers over the car and then, suddenly, daffodil in hand, she races off down the street yelling as she goes that she will meet us at the South Pole. (She is later apprehended by the authorities wrestling with a mannequin in a shop window.) (Malpas, 1992, p.59)

Notice how significantly this case differs from the example that I borrowed from Wittgenstein in Chapter 3, §V. (See PI §207.) And recall Cavell's suggestion (noted in Chapter 1, §V) that scepticism is an expression of our fear of madness. See Cavell (1990), p.85f.

[90] The term is Campbell's. See Campbell (1979), pp.201-212. Campbell makes use of the notion of a true or real self, and one's self-identifying desires are correspondingly one's true or real desires. However, his use of these notions seems self-consciously normative, rather than metaphysical, so that there is no substantive conflict between his account and my own. As a matter of terminological preference I shall think of Campbell's "real self" simply as the self that I want to be, but which I may or may not yet be.

self-description that is false. I *have* those desires of which I long to be rid, but they are not part of my ideal image of myself.[91] This gravitation to an ideal may be just what I need to overcome my unsavoury urges. "I am not that sort of a person," I tell myself, and by so insisting I can sometimes change myself so that I no longer have the desires from which I sought relief, or at least, so that those desires no longer pull me with the same force. A false description has, partly through repeated and forceful affirmation, become a true one--or at least less of a distortion than it was. Adopting a false self-description, I have helped to liberate myself, but in the process the applicability of my self-description has also changed.

What if it had not? What if my distorted self-image did nothing to rid me of my compulsive desires? In such a case we are faced with the sort of covertly coercive false unity of the self against which Lugones warns. In such a case I embrace a faulty self-description that serves to mask an underlying threat to my agency, reinforcing its efficacy by disguising its coercive nature.

Need this always be so? Imagine that I wrongly believe that I can stop smoking at any time, but rationalize my smoking by saying that, all things considered, I simply *prefer* to smoke. In some instances such self-deception (or mere self-ignorance) might be an advantage for my agency. Maybe I lead a stressful life without the chance to deal "properly" with my stress, and smoking helps me to cope. Trying to quit might be too much for me; it might send me into nervous collapse.

It may be appropriate to say here that a faulty self-description enhances my capacity for agency. But it does so under circumstances that present other threats to my agency. My self-ignorance helps here, because it lets me deal with other pressures in my life, pressures that might overwhelm me but for my distorted self-concept. This does not show that faulty descriptions can be of benefit to my agency, only if they stand opposed to other threats to my agency. Nor shall I try to find examples or arguments that would entail such a consequence. All that I wish to claim is that self-knowledge is very often agency-enhancing. That it is not always so can be no objection.[92]

A more interesting question in the present context, then, is whether an individual's self-descriptions are desirable. Clearly, this issue must be settled in specific cases. As such, it is hard to detach from two more familiar questions: "How should I live my life?" and "What kinds of social order are desirable?" The former pertains to the kinds of self-descriptions I should adopt from among those available to me. The latter pertains to what ways of articulating my attitudes *are* so available to me. If we can appeal to the desirability of a capacity for agency, then certain kinds of self-descriptions do seem to be open to criticism. Certain *disunities* of the self plainly interfere with one's capacity to make rational assessments of one's situation and

---

[91] See Noddings (1984), pp.49-51.

[92] It also seems plausible that the case of the deluded smoker is one in which a false self-description provides the means for coping, not just with pre-existing threats to one's agency, but with pre-existing false self-descriptions, incurred by those very threats, but also reinforcing them. (The relation would, again, be a dialectical one.) Whether this were so would have to be established in particular cases, but I regard it as a likely finding in many cases.

then to act on those assessments. The examples of self-division under oppression and multiple-personalities bear this out. And certain false unities--self-descriptions that mask debilitating self-divisions--seem likewise undesirable from this perspective. My suggestion might be summarized by saying that, very often, correct self-descriptions are desirable, and that where they are not, some false self-predications may still be preferable to others. Perhaps we can, in turn, judge a society on the basis of the repertoire of articulations that it makes available to its citizens, especially to its least well-off. But here is not the place to examine these problems. That some sort of self-unity is presupposed by linguistic behaviour is enough for my ends.

VII. Concluding Therapeutic Remarks

My criticisms of metaphysical realism and of relativism, of the dichotomy of objectivity and subjectivity, have been couched in the rhetoric of mental illness. Epistemic neurosis, I have maintained, is the natural philosophical malady of those who indulge in the vice of dichotomizing the objective and the subjective. To deal with this malady I have proposed a course of philosophical therapy that tries to preserve the virtues of realism and relativism without falling back into psycho-philosophical disorder. However, it might be protested that the rhetoric of mental instability--like the rhetoric of war, which I have tried to avoid in my critique--is *cheap*. It is a favourite ploy of the modern era to dismiss those whose views we do not like as "crazy," just as it was a favourite and more dangerous ploy in older times to condemn people as "heretics" and "witches."

I think that my use of the term 'neurosis' has been reasonable, since I have tried in its application to criticize philosophical positions for entertaining theses--especially sceptical theses--that they cannot afford to take seriously, if their own positive proposals are to have a chance of cognitive success. I have, moreover, tried to show in some detail how these epistemic neuroses reveal themselves in the views I have examined, and I have suggested ways of avoiding the temptations that they present. That I have indulged in rhetoric and that my rhetoric is designed to persuade I do not deny. But it is part of my position that no clean break can be found between logic and rhetoric. The idea of a pristine language, appearing as an unfettered manifestation of "pure reason," is just the idea of a language whose meanings transcend our linguistic and epistemic capacities. On the other hand, my rejection of non-normative accounts of meaning does not give me license to discard the idea of good reasoning. I do not think that my arguments (or my rhetoric) depend on doing so.

I have been trying to make a case for thinking that influential strands of modern epistemology and philosophy of language suffer from a kind of disorder, analogous to an undesirable disunity of the self. But the metaphor of neurosis that I have used is a classification no longer employed by psychiatrists, psychologists and therapists.[93] Were

---

[93] Since 1980 the American Psychiatric Association has recognized a number of "dissociative disorders," including such maladies as multiple-personality disorder, psychogenic amnesia, psychogenic fugue and depersonalization disorder. These are all disorders which would formerly have been classified as "neuroses," since their sufferers retain some grip on reality and can be conscious of their troubles,

I to search for a successor-metaphor, maybe I could find something in the lexicon of psychiatric disorders that would fit the case even more suggestively than "neurosis," or perhaps I could find a way of distinguishing a number of different epistemic disorders--or, perhaps no successor would bear the weight that I have placed on "neurosis."

Do philosophers suffer from an epistemic dissociative disorder? In Joan Frances Casey's autobiography the tale told by one of her former personalities alternates with the journal-entries of her therapist, Lynn Wilson. Wilson describes the paradoxical state of Casey's psychic affairs after a year and a half of treatment:

> *I am continually struck by the essential contradiction in Jo Casey. She is both terribly fragmented and tremendously functional. When I first heard the Kendra personality (who rescues Renee when she gets in over her head) say, "We can do anything!," I considered it bravado and exaggeration. Now I see that, in comparison with most people, Kendra is right. The system of personalities can, by most criteria, do anything they attempt to do, as long as no one personality is expected to carry on for an extended period.*[94]

When Hume leaves his study, he tells us, his sceptical worries vanish, pushed out by the forces of habit and custom:

> I dine, I play a game of back-gammon, I converse, and am merry with my friends; and when after three or four hour's amusement, I wou'd return to these speculations, they appear so cold, and strain'd, and ridiculous, that I cannot find in my heart to enter into them any farther. (THN 269)

He leads a double existence, dictated in a way by the "opinion" of the same name, which--philosophically--is "only a palliative remedy" (THN 211). But this is not just any double existence; it is an intractable one, if we take seriously the two selves that are in conflict. The sceptical self cannot acknowledge the merry-maker, and the merry-maker cannot take the sceptic seriously. But maybe this is not so different from other cases of dissociation or self-division. Overcoming division requires an awareness of one's plurality, and this awareness seems tenuous among some philosophers, more developed among others who have remembered Hume. But Hume and his followers must shrug their shoulders and resign themselves to such division. Integration would dictate abandoning assumptions at the heart of their self-interpretations: the dichotomy of objectivity and subjectivity and the non-normative nature of meaning and truth. Such abandonment is the therapeutic course that I have proposed, and it means becoming a different kind of philosopher with a different set of self-descriptions.

I conclude with another passage borrowed from Casey's compelling story, a passage whose rhetorical utility in the present context is hardly to be believed. Would that I could put it to better use. But if my arguments have been convincing, maybe a mere flourish will suffice:

---

[93] though unable to overcome them without help. See American Psychiatric Association (1980), pp.253-260.

[94] Casey and Wilson (1991), p.125.

Jo felt let down by her mind, by her extraordinary analytic abilities. She had always thought she was smart and had approached any intellectual challenge with glee. When presented with the philosophical hypothesis that people could be nothing more than minds in a vat, hallucinating reality, Jo wasn't perturbed or perplexed. Amnesia and the familiar feeling of "I just got dropped in here somehow" enabled Jo to see how this improbable hypothesis served as an analogy for her life. She figured that she could be a very contented mind-in-a-vat, but Lynn had forced her to accept she wasn't alone in a vat or in her body. This was a problem that she could not puzzle.[95]

---

[95] *Ibid.*, p.85.

Bibliography

Adams, Douglas. 1979. *The Hitch-hiker's Guide to the Galaxy.* London: Pan Books Ltd.

Allison, Henry. 1983. *Kant's Transcendental Idealism.* New Haven: Yale University Press.

Allison, Ralph B. 1977. "Introduction." In Hawksworth and Schwarz (1977), pp.v-viii.

Alston, William P. 1978-79. "Yes, Virginia, There Is a Real World." *Proceedings and Addresses of the American Philosophical Association* 52: 779-808.

Althusser, Louis. 1969. *For Marx.* Trans. Ben Brewster. New York: Pantheon Books.

_____. 1971. *Lenin and Philosophy and Other Essays.* Trans. Ben Brewster. New York: Monthly Review Press.

American Psychiatric Association. 1980. *Diagnostic and Statistical Manual of Mental Disorders.* 3rd. ed. Washington: American Psychiatric Association.

Anzaldúa, Gloria. 1987. *Borderlands/La Frontera: The New Mestiza.* San Francisco: Spinsters/Aunt Lute.

Appiah, Anthony. 1985. "The Uncompleted Argument: DuBois and the Illusion of Race." *Critical Inquiry* 12: 21-37.

Arbib, Michael and Mary Hesse. 1986. *The Construction of Reality.* Cambridge: Cambridge University Press.

Austin, J. L. 1962. *How to Do Things with Words.* Ed. J. O. Urmson. New York: Oxford University Press.

Ayer, A. J. 1952. *Language, Truth and Logic.* 2nd ed. New York: Dover Publications, Inc.

Baier, Annette. 1985. *Postures of the Mind: Essays on Mind and Morals.* Minneapolis: University of Minnesota Press.

_____. 1986. "Trust and Antitrust." *Ethics* 96: 231-260.

Baker, G. P. and P. M. S. Hacker. 1984a. *Scepticism, Rules and Language.* Oxford: Basil Blackwell.

_____. 1984b. *Language, Sense and Nonsense.* Oxford: Basil Blackwell.

_____. 1990. "Malcolm on Language and Rules." *Philosophy* 65: 167-179.

Barnes, Barry. 1987. "Concept Application as Social Activity." *Crítica* 19, no.56: 19-44.

Barnes, Barry and David Bloor. 1982. "Relativism, Rationalism and the Sociology of Knowledge." In Hollis and Lukes (1982), pp.21-47.

Bartky, Sandra. 1982. "Narcissism, Femininity and Alienation." *Social Theory and Practice* 8: 127-144.

Baynes, Kenneth, James Bohman and Thomas McCarthy, eds. 1987. *Philosophy: End or Transformation*? Cambridge, Ma.: MIT Press.

Benaceraff, Paul. 1985. "Skolem and the Sceptic." *Aristotelian Society Supplementary Volume* 58: 85-115.

Bennett, Jonathan. 1966. *Kant's Analytic*. Cambridge: Cambridge University Press.

Benton, Ted. 1984. *The Rise and Fall of Structural Marxism: Althusser and his Influence*. London: MacMillan.

Berkeley, George. 1982. *A Treatise Concerning the Principles of Human Knowledge*. Ed. Kenneth Winkler. Indianapolis: Hackett Publishing Company.

Bilgrami, Akeel. 1992. "Self-Knowledge and Resentment." Forthcoming.

Blackburn, Simon. 1984a. *Spreading the Word: Groundings in the Philosophy of Language*. Oxford: Clarendon Press.

_____. 1984b. "The Options in Debating Realism." *Dalhousie Review* 64: 553-572.

Bloor, David. 1976. *Knowledge and Social Imagery*. London: Routledge and Kegan Paul.

_____. 1983. *Wittgenstein: A Social Theory of Knowledge*. London: Macmillan.

BonJour, Laurence. 1985. *The Structure of Empirical Knowledge*. Cambridge, Mass.: Harvard University Press.

Bordo, Susan. "The Cartesian Masculinization of Thought." *Signs* 11 (1986): 439-456.

Brennan, Teresa. 1993. "The Labour Theory of Value without the Subject/Object Distinction." Distinguished Visitor's Lecture, University of Alberta, Edmonton, 24 March.

Brittan, Gordon. 1978. *Kant's Theory of Science*. Princeton: Princeton University Press.

Brueckner, Anthony. 1986. "Brains in a Vat." *The Journal of Philosophy* 83: 148-167.

Burge, Tyler. 1979. "Individualism and the Mental." *Midwest Studies in Philosophy* 4: 73-121.

_____. 1986. "Cartesian Error and the Objectivity of Perception." In *Subject, Thought and Context.* Ed. Philip Pettit and John McDowell. Oxford: Clarendon Press, 1986, pp.117-136.

_____. 1988. "Individualism and Self-knowledge." *The Journal of Philosophy* 85: 649-663.

Burns, Steven A. M. 1989. "The Place of Art in a Reasonable Education." In *Reason in Teaching and Education.* Ed. William Hare. Halifax: School of Education, Dalhousie University, pp.23-40.

Campbell, Richmond. 1979. *Self-Love and Self-Respect: A Philosophical Study of Egoism.* Ottawa: Canadian Library of Philosophy.

_____. 1993. "The Virtues of Feminist Empiricism." ts. Department of Philosophy, Dalhousie University.

Carnap, Rudolf. 1956. *Meaning and Necessity.* 2nd ed. Chicago: University of Chicago Press.

_____. 1963. "Intellectual Autobiography." In *The Philosophy of Rudolf Carnap.* Ed. P. A. Schilpp. La Salle: Open Court, 1963, pp.3-84.

Casey, Joan Frances with Lynn Wilson. 1991. *The Flock.* New York: Alfred A. Knopf.

Cavell, Stanley. 1990. "The Argument of the Ordinary: Scenes of Instruction in Wittgenstein and Kripke." In *Conditions Handsome and Unhandsome: The Constitution of Emersonian Perfectionism.* Chicago: University of Chicago Press, 1990.

Chambers, Ross. 1992. "No Montagues without Capulets: Some Thoughts on 'Cultural Identity'." ts. Department of Romance Languages and Literatures, University of Michigan, Ann Arbor.

Chisholm, Roderick. 1989. *Theory of Knowledge.* 3rd ed. Englewood Cliffs, N. J.: Prentice Hall.

Chomsky, Noam. 1989. *Necessary Illusions: Thought Control in Democratic Societies.* Toronto: CBC Enterprises.

_____. 1993. *Year 501: The Conquest Continues.* Montreal: Black Rose Books.

Cixous, Hélène. 1976. "The Laugh of the Medusa." Trans. Keith Cohen and Paula Cohen. *Signs* 1: 875-893.

Code, Lorraine. 1991. *What Can She Know?* Ithaca: Cornell University Press.

Cohen, Gerald A. 1978. *Karl Marx's Theory of History: A Defence.* Princeton: Princeton University Press.

Collingwood, R. G. 1940. *An Essay on Metaphysics.* Oxford: Clarendon Press.

Coppock, Paul. 1987. "Putnam's Transcendental Argument." *Pacific Philosophical Quarterly* 68: 14-28.

Curley, E. M. 1978. *Descartes Against the Skeptics*. Cambridge, Ma.: Harvard University Press.

D'Amico, Robert. 1989. *Historicism and Knowledge*. New York: Routledge.

Davidson, Donald. 1970. "Mental Events." In *Experience and Theory*. Eds. L. Foster and J. W. Swanson. Amherst: University of Massachusetts Press, 1970, pp.79-101.

_____. 1984a. *Inquiries into Truth and Interpretation*. Oxford: Clarendon Press.

_____. 1984b. "First Person Authority." *Dialectica* 38: 101-111.

_____. 1985. "Deception and Division." In *Actions and Events: Perspectives on the Philosophy of Donald Davidson*. Ed. Ernest LePore and Brian P. McLaughlin. Oxford: Basil Blackwell, 1985, pp.138-148.

_____. 1986. "A Nice Derangement of Epitaphs." In LePore (1986), pp.433-446.

_____. 1987. "Knowing One's Own Mind." *Proceedings and Addresses of the American Philosophical Association* 60: 441-458.

_____. 1989. "The Myth of the Subjective." In *Relativism: Interpretation and Confrontation*. Ed. Michael Krausz. Notre Dame: University of Notre Dame Press, 1989, pp.159-172.

Derrida, Jacques. 1988. *Limited Inc*. Ed. Gerald Graff. Evanston: Northwestern University Press.

Descartes, René. 1984. *The Philosophical Writings of Descartes*, 2 vols. Trans. John Cottingham, Robert Stoothoff and Dugald Murdoch. Cambridge: Cambridge University Press.

Devitt, Michael. 1981. *Designation*. New York: Columbia University Press.

_____. 1984. *Realism and Truth*. Oxford: Basil Blackwell.

Devitt, Michael and Kim Sterelny. 1987. *Language and Reality: An Introduction to the Philosophy of Language*. Oxford: Basil Blackwell.

DuBois, W. E. B. 1968. *The Autobiography of W. E. B. DuBois: A Soliloquy on Viewing My Life from the Last Decade of Its First Century*. Ed. Herbert Aptheker. N.p.: International Publishers.

Dummett, Michael. 1975. "What Is a Theory of Meaning?" In Guttenplan (1975), pp.97-138.

_____. 1976. "What Is a Theory of Meaning? II" In *Truth and Meaning: Essays in Semantics*. Eds. Gareth Evans and John McDowell. Oxford: Clarendon Press, 1976, pp.67-137.

_____. 1979. "Realism." In *Truth and Other Enigmas*. Cambridge: Cambridge University Press, 1979.

Eagleton, Terry. 1991. *Ideology: An Introduction*. London: Verso.

Ehrenreich, Barbara and Deirdre English. 1978. *For Her Own Good: 150 Years of the Experts' Advice to Women*. New York: Anchor Books.

Ellison, Ralph. 1989. *Invisible Man*. New York: Vintage.

Feyerabend, Paul. 1987. *Farewell to Reason*. London: Verso.

_____. 1988. *Against Method*. rev. ed. London: Verso.

Field, Hartry. 1972. "Tarski's Theory of Truth." *Journal of Philosophy* 69: 347-375.

_____. 1982. "Realism and Relativism." *Journal of Philosophy* 79: 553-567.

Fogelin, Robert. 1976. *Wittgenstein*. London: Routledge and Kegan Paul.

Fraser, Nancy. 1989. *Unruly Practices: Power, Discourse and Gender in Contemporary Social Theory*. Minneapolis: University of Minnesota Press.

Frye, Marilyn. 1983. *The Politics of Reality: Essays in Feminist Theory*. Trumansburg, N.Y.: The Crossing Press.

Gauthier, David. 1986. *Morals by Agreement*. Oxford: Clarendon Press.

Giancarlo, Thomas J. 1991. "Multiple Personality Disorder: A Challenge to Practitioners." *Families in Society: The Journal of Contemporary Human Services* 72: 95-102.

Godlovitch, Stan. 1992. Philosophy Club Lecture. University of Alberta, Edmonton.

Goldman, Alvin. 1975. "Innate Knowledge." In *Innate Ideas*. Ed. Stephen P. Stich. Berkeley: University of California Press, 1975, pp.111-120.

_____. 1986. *Epistemology and Cognition*. Cambridge, Ma.: Harvard University Press.

_____. 1992. *Liaisons: Philosophy Meets the Cognitive and Social Sciences*. Cambridge, Ma.: MIT Press.

Grandy, Richard. 1973. "Reference, Meaning and Belief." *Journal of Philosophy* 70: 439-452.

Grimshaw, Jean. 1986. *Philosophy and Feminist Thinking*. Minneapolis: University of Minnesota Press. Also published as *Feminist Philosophers: Women's Perspectives on Philosophical Traditions*. Brighton: Wheatsheaf, 1986.

Gross, Michael and Mary Beth Averill. 1983. "Evolution and Patriarchal Myths of Scarcity and Competition." In Harding and Hintikka (1983), pp.71-95.

Gueroult, Martial. 1984. *Descartes' Philosophy Interpreted According to the Order of Reasons, Volume I: The Soul and God*. Trans. Roger Ariew. Minneapolis: University of Minnesota Press.

Guttenplan, Samuel ed. 1975. *Mind and Language*. Oxford: Clarendon Press.

Habermas, Jürgen. *Toward a Rational Society*. 1970a. Trans. Jeremy J. Shapiro. Boston: Beacon Press.

_____. 1970b. "On Systematically Distorted Communication." *Inquiry* 13: 205-218.

_____. 1971. *Knowledge and Human Interests*. Trans. Jeremy J. Shapiro, Boston: Beacon Press.

_____. 1987. "Philosophy as Stand-In and Interpreter." Trans. Christian Lenhardt. In Baynes, Bohman and McCarthy (1987), pp.296-315.

Hacker, P. M. S. 1992. "Malcolm and Searle on 'Intentional Mental States'." *Philosophical Investigations* 15: 245-275.

Hacking, Ian. 1982. "Language, Truth and Reason." In Hollis and Lukes (1982), pp.48-66.

_____. 1983. *Representing and Intervening: Introductory Topics in the Philosophy of Natural Science*. Cambridge: Cambridge University Press.

_____. 1986. "The Invention of Split Personalities: An illustration of Michel Foucault's doctrine of the constitution of the subject." In *Human Nature and Natural Knowledge*. Eds. Alan Donagan, Anthony N. Perovich, Jr. and Michael V. Wedin. Dordrecht: D. Reidel, 1986, pp.63-85.

_____. 1991. "Two Souls in One Body." *Critical Inquiry* 17: 838-867.

Hahn, Hans, Otto Neurath and Rudolf Carnap. 1973. "The Scientific Conception of the World: The Vienna Circle." Trans. Paul Foulkes and Marie Neurath. In *Otto Neurath: Empiricism and Sociology*. Eds. Marie Neurath and Robert S. Cohen. Dordrecht: D. Reidel, pp.299-318.

Hanen, Marsha. "Toward Integration." In Hanen and Nielsen (1987), pp.1-21.

Hanen, Marsha and Kai Nielsen, eds. 1987. *Science, Morality and Feminist Theory. Canadian Journal of Philosophy, Supplementary Volume 13*.

Harding, Sandra and Merrill B. Hintikka eds. 1983. *Discovering Reality: Feminist Perspectives on Epistemology, Metaphysics, Methodology, and Philosophy of Science*. Dordrecht: D. Reidel.

Harman, Gilbert. 1965. "The Inference to the Best Explanation." *The Philosophical Review* 74: 88-95.

Haugeland, John. 1982. "Weak Supervenience." *American Philosophical Quarterly* 19: 93-103.

Hawksworth, Henry and Ted Schwarz. 1977. *The Five of Me*. Chicago: Henry Regnery Company.

Heldke, Lisa. 1988. "Recipes for Theory Making." *Hypatia* 3: 15-29.

Hesse, Mary. 1980. *Revolutions and Reconstructions in the Philosophy of Science*. Bloomington: Indiana University Press.

Hockney, Donald. 1975. "The Bifurcation of Scientific Theories and Indeterminacy of Translation." *Philosophy of Science* 42: 411-427.

Hollis, Martin and Steven Lukes eds. 1982. *Rationality and Relativism*. Cambridge, Ma.: MIT Press.

Horkheimer, Max. 1982. *Critical Theory: Selected Essays*. New York, Continuum.

Hubbard, Ruth. 1983. "Have Only Men Evolved?" In Harding and Hintikka (1983), pp.45-69.

Hume, David. 1978. *A Treatise of Human Nature*. Ed. L. A. Selby-Bigge. Oxford: Clarendon Press.

Hunter, Bruce. 1982. "Critical Notice of Richard Rorty, *Philosophy and the Mirror of Nature*." *Canadian Journal of Philosophy* 12: 621-645.

Hymers, Michael. 1989a. "Bad Faith." *Philosophy* 64: 397-402.

_____. 1989b. "Putnam, Kant and 'Real' Possibility." Proceedings of the Canadian Graduate Students' Symposium in Philosophy, Memorial University of Newfoundland. Forthcoming.

_____. 1991. "Something Less than Paradise: The Magic of Modal Realism." *Australasian Journal of Philosophy* 69: 251-263.

_____. 1992. "The Structure of Kant's Refutation of Idealism." ts. University of Alberta.

_____. 1993. "Review of J. E. Malpas, *Donald Davidson and the Mirror of Meaning*." *Canadian Philosophical Reviews* 13, no.3 (1993). Forthcoming.

Janik, Allan. 1985. "Wittgenstein, Marx and Sociology." In *Essays on Wittgenstein and Weininger*. Amsterdam: Rodopi, 1985, pp.136-157.

Kant, Immanuel. 1923. *Kritik der reinen Vernunft*. 6th ed. rev. Ed. Benno Erdmann. Berlin: Walter de Gruyter & Co.

_____. 1965. *Critique of Pure Reason*. Trans. Norman Kemp Smith. New York: St. Martin's Press.

Kenny, Anthony. 1987. *Descartes: A Study of His Philosophy*. 2nd. ed. New York: Garland Publishing, Inc.

Kinghan, Michael. 1986. "The External World Sceptic Escapes Again." *Philosophia* 16: 161-166.

Kripke, Saul. 1980. *Naming and* Necessity. Cambridge, Ma.: Harvard University Press.

_____. 1982. *Wittgenstein on Rules and Private Language*. Cambridge, Ma.: Harvard University Press.

Kuhn, Thomas. 1969. *The Structure of Scientific Revolutions*. 2nd ed. Chicago: University of Chicago Press.

Lacan, Jacques. 1989. "The Mirror Stage as Formative of the Function of the I as Revealed in Psychoanalytic Experience." Trans. Alan Sheridan. In *Contemporary Critical Theory*. Ed. Dan Latimer. San Diego: Harcourt Brace Jovanovich, 1989, pp.502-509.

"Lagos men fear rumour: mob attacks spurred by tales of genital thefts." *Calgary Herald*, 30 October 1990, p.F12.

Lenin, V. I. 1927. *Materialism and Empirio-Criticism*. Trans. David Kvitko. Collected Works XIII. Ed. Alexander Trachtenberg. London: Martin Lawrence, Ltd.

LePore, Ernest, ed. 1986. *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson*. Oxford: Basil Blackwell.

LePore, Ernest and Barry Loewer. 1988. "A Putnam's Progress." *Midwest Studies in Philosophy* 12: 459-473.

Lewis, David. 1983. "New Work for a Theory of Universals." *Australasian Journal of Philosophy* 61: 343-377.

_____. 1984. "Putnam's Paradox." *Australasian Journal of Philosophy* 62 (1984): 221-236.

_____. 1987. *On the Plurality of Worlds*. Oxford: Basil Blackwell, Ltd.

Lloyd, Genevieve. 1984. *The Man of Reason: Male and Female in Western Philosophy*. London: Methuen.

Locke, John. 1964. *An Essay concerning Human Understanding*. Ed. A. D. Woozley. New York: New American Library.

Lovibond, Sabina. 1983. *Realism and Imagination in Ethics*. Oxford: Basil Blackwell.

Lugones, Maria C. 1990. "Structure/Antistructure and Agency under Oppression." *Journal of Philosophy* 87: 500-507.

Lukes, Steven. 1982. "Relativism in Its Place." In Hollis and Lukes (1982), pp.261-305.

MacIntyre, Alasdair. 1970a. "Is Understanding Religion Compatible with Believing?" In Wilson (1970), pp.62-77.

_____. 1970b. "The Idea of a Social Science." In Wilson (1970), pp.112-130.

_____. 1987. "Relativism, Power, and Philosophy." In Baynes, Bohman and McCarthy (1987), pp.385-411.

Mackie, J. L. 1977. *Ethics: Inventing Right and Wrong.* Harmondsworth: Penguin Books Ltd.

MacKinnon, Catharine A. 1982. "Feminism, Marxism, Method, and the State: An Agenda for Theory." *Signs* 7: 515-544.

_____. 1987. *Feminism Unmodified: Discourses on Life and Law.* Cambridge, Ma.: Harvard University Press.

Malcolm, Norman. 1986. *Nothing Is Hidden.* Oxford: Basil Blackwell.

_____. 1989. "Wittgenstein on Language and Rules." *Philosophy* 64: 5-28.

Malpas, J. E. 1992. *Donald Davidson and the Mirror of Meaning.* Cambridge: Cambridge University Press.

Mannheim, Karl. [1936]. *Ideology and Utopia.* Trans. Louis Wirth and Edward Shils. New York: Harcourt Brace Jovanovich, n.d.

Marcuse, Herbert. 1962. *Eros and Civilization: A Philosophical Inquiry into Freud.* 2nd. ed. New York: Vintage Books.

_____. 1964. *One-Dimensional Man.* Boston: Beacon Press.

Martin, Biddy and Chandra Talpade Mohanty. 1986. "Feminist Politics: What's Home Got to Do with It?" In *Feminist Studies/Critical Studies.* Ed. Teresa de Lauretis. Bloomington: Indiana University Press, 1986, pp.191-212.

Martin, Robert M. 1992. *There Are Two Errors in the the Title of This Book.* Peterborough: Broadview Press.

Marx, Karl. 1953. *Grundrisse der Kritik der politischen Ökonomie.* Berlin: Dietz Verlag.

_____. 1967. *Capital.* 3 vols. Trans. Samuel Moore and Edward Aveling. Ed. Friedrich Engels. New York: International Publishers.

_____. 1977. *Economic and Philosophic Manuscripts of 1844.* Trans. ? Moscow: Progress Publishers.

Marx, Karl and Friedrich Engels. 1970. *The German Ideology, Part I.* Trans. R. Pascal. Ed. C. J. Arthur. New York: International Publishers.

_____. 1969. *Karl Marx und Friedrich Engels--Werke*, 42 vols. Berlin: Dietz Verlag.

_____. 1977. Karl Marx: Selected Writings. Ed. David McLellan. Oxford: Oxford University Press.

Merskey, Harold. 1993. "Peer Review." *The Sciences* 33, no.2, p.4.

Moore, G. E. 1959. *Principia Ethica*. Cambridge: Cambridge University Press.

Mullett, Sheila. 1987. "Only Connect: The Place of Self-Knowledge in Ethics." In Hanen and Nielsen (1987), pp.309-338.

Nagel, Thomas. 1986. *The View from Nowhere*. Oxford: Oxford University Press.

Neurath, Otto. 1983. *Otto Neurath--Philosophical Papers, 1913-1946.* Trans. and eds. Robert S. Cohen and Marie Neurath. Dordrecht: D. Reidel Publishing Company.

Nielsen, Kai. 1987. "Feminist Theory--Some Twistings and Turnings." In Hanen and Nielsen (1987), pp.383-418.

_____. 1988. "Farewell to the Tradition: Doing without Metaphysics and Epistemology." ts. University of Calgary.

Nietzsche, Friedrich. 1966. *Beyond Good and Evil.* Trans. Walter Kaufmann. New York: Vintage Books.

"Nigerians fear dismembering as tales of genital theft grow." *Globe and Mail*, 30 October 1990, p.A14.

Noddings, Nell. *Caring: A Feminine Approach to Ethics and Moral Education*. Berkeley: University of California Press, 1984.

Noel, Justine J. 1989. "The Appeal to Realism: Towards a Non-Anthropocentric Science." Proceedings of the Canadian Graduate Students' Symposium in Philosophy, Memorial University of Newfoundland. Forthcoming.

Overall, Christine. 1988. "Feminism, Ontology, and 'Other Minds'." In *Feminist Perspectives: Philosophical Essays on Method and Morals.* Eds. Lorraine Code, Sheila Mullett and Christine Overall. Toronto: University of Toronto Press, 1988, pp.89-106.

Parekh, Bhiku. 1982. *Marx's Theory of Ideology*. Baltimore: Johns Hopkins Press.

Parfit, Derek. 1984. *Reasons and Persons*. Oxford: Clarendon Press.

Penelhum, Terence. 1975. *Hume*. London: The MacMillan Press Ltd.

Plato. *Plato's Republic*. Trans. G. M. A. Grube. Indianapolis: Hackett Publishing Company, 1974.

Pollock, John. 1986. *Contemporary Theories of Knowledge*. Totowa, N.J.: Rowman and Littlefield.

Putnam, Frank W. 1993. "Peer Review." *The Sciences* 33, no.2, p.4.

Putnam, Hilary. 1975. *Mind, Language and Reality: Philosophical Papers, Volume 2*. Cambridge: Cambridge University Press.

_____. 1978. *Meaning and the Moral Sciences*. London: Routledge & Kegan Paul.

_____. 1981. *Reason, Truth and History*. Cambridge: Cambridge University Press.

_____. 1983. *Realism and Reason: Philosophical Papers, Volume 3*. Cambridge: Cambridge University Press.

_____. 1987. *The Many Faces of Realism*. LaSalle, Illinois: Open Court.

_____. 1988. *Representation and Reality*. Cambridge, Ma.: MIT Press.

_____. 1990. *Realism with a Human Face*. Ed. James Conant. Cambridge, Ma.: Harvard University Press.

_____. 1992a. *Renewing Philosophy*. Cambridge, Ma.: Harvard University Press.

_____. 1992b. "Truth, Activation Vectors and Possession of Conditions for Concepts." *Philosophy and Phenomenological Research* 52: 431-447.

Quine, W. V. 1960. *Word and Object*. Cambridge, Ma.: MIT Press.

_____. 1969. *Ontological Relativity and Other Essays*. New York: Columbia University Press.

_____. 1974. *The Roots of Reference*. LaSalle, Illinois: Open Court.

_____. 1975. "The Nature of Natural Knowledge." In Guttenplan (1975), pp.67-81.

_____. 1979. "A Postscript on Metaphor." In *On Metaphor*. Ed. Sheldon Sacks. Chicago: University of Chicago Press, 1979, pp.159-160.

_____. 1980. *From a Logical Point of View*. rev. ed. Cambridge, Ma.: Harvard University Press.

_____. 1990. *Pursuit of Truth*. Cambridge, Ma.: Harvard University Press.

Rawls, John. 1971. *A Theory of Justice*. Cambridge, Ma.: Belknap Press, 1971.

Reynolds, Steven L. 1992. "Descartes and First Person Authority." *History of Philosophy Quarterly* 9: 181-189.

Root, Michael. 1986. "Davidson and Social Science." In LePore (1986), pp.272-304.

Rorty, Richard. 1979. *Philosophy and the Mirror of Nature*. Princeton: Princeton University Press.

_____. 1989. Contingency, Irony and Solidarity. Cambridge: Cambridge University Press.

_____. 1991a. *Objectivity, Relativism and Truth: Philosophical Papers, Volume 1*. Cambridge: Cambridge University Press.

_____. 1991b. "Feminism and Pragmatism." *Radical Philosophy* 59 Autumn: 3-14.

_____. 1992a. "Putnam on Truth." *Philosophy and Phenomenological Research* 52: 415-418.

_____. 1992b. "'We Anti-representationalists'." *Radical Philosophy* 60 Spring: 40-42.

Rose, Jacqueline. 1982. "Introduction II." In Jacques Lacan. *Feminine Sexuality*. Trans. Jacqueline Rose. Eds. Juliet Mitchell and Jacqueline Rose. New York: W. W. Norton & Company, 1982, pp.27-57.

Rose, Steven, R. C. Lewontin and Leon J. Kamin. 1984. *Not in Our Genes: Biology, Ideology and Human Nature*. Harmondsworth: Penguin Books.

Rosen, Michael. 1983. "Critical Theory: Between Ideology and Philosophy." In *The Need for Interpretation*. Eds. Sollace Mitchell and Michael Rosen. London: The Athalone Press, 1983, pp.90-117.

Russell, Bertrand. 1919. "Descriptions." In *Introduction to Mathematical Philosophy*. London: George Allen and Unwin Ltd., pp.167-80.

Ryle, Gilbert. 1949. *The Concept of Mind*. New York: Barnes & Noble, Inc.

Sankey, Howard. 1990. "In Defence of Untranslatability." *Australasian Journal of Philosophy* 68: 1-21.

Sartre, Jean-Paul. 1956. *Being and Nothingness*. Trans. Hazel Barnes. New York: Washington Square Press.

Scheman, Naomi. 1983. "Individualism and the Objects of Psychology." In Harding and Hintikka (1983), pp.225-244.

Schouls, P. A. 1980. *The Imposition of Method: A Study of Descartes and Locke*. Oxford: Clarendon Press.

_____. 1993. "Human Nature and the Argument in Descartes' *Meditations*." In *Reason, Will and Sensation: Studies in Cartesian Metaphysics*. Ed. John Cottingham. Oxford: Oxford University Press, 1993, forthcoming.

Seller, Anne. 1988. "Realism versus Relativism: Towards a Politically Adequate Epistemology." In *Feminist Perspectives in Philosophy*. Eds. Morwenna Griffiths and Margaret Whitford. Bloomington: Indiana University Press, 1988, pp.169-186.

"Sex-organ scare touches off rioting." *Edmonton Journal* 13 November 1990, p.B4.

Sperber, Dan. 1982. "Apparently Irrational Beliefs." In Hollis and Lukes (1982), pp.149-180.

Staten, Henry. 1984. *Wittgenstein and Derrida*. Lincoln: University of Nebraska Press.

Sterelny, Kim. 1991. "Confessions of a Causal Theorist." Philosophy Club Lecture. University of Alberta, Edmonton.

Stroud, Barry. 1984. *The Significance of Philosophical Scepticism*. Oxford: Clarendon Press.

Sturgeon, Nicholas. 1985. "Moral Explanations." In *Morality, Reason and Truth*. Eds. David Copp and David Zimmerman. Totowa: Rowman and Allanheld, 1985, pp.49-78.

_____. 1986. "Harman on Moral Explanations of Natural Facts." In *Spindel Conference 1986: Moral Realism*. Ed. Norman Gillespie. Memphis: *The Southern Journal of Philosophy, Supplementary Volume XXIV*, 1986, pp.69-78.

Szabados, Béla. 1981a. "Wittgenstein on 'Mistrusting One's Own Belief'." *Canadian Journal of Philosophy* 11: 603-612.

_____. 1981b. "Wittgenstein on Belief." *Philosophical Papers* 10: 24-34.

Taylor, Charles. 1982. "Rationality." In Hollis and Lukes (1982), pp.87-105.

_____. 1985. *Human Agency and Language: Philosophical Papers, Volume 1*. Cambridge: Cambridge University Press.

Tichy, Pavel. 1986. "Putnam on Brains in a Vat." *Philosophia* 16: 137-146.

"Vanishing organs spark mob attacks." *Montreal Gazette*, 4 November 1990, p.B4.

West, Cornel. 1989. *The American Evasion of Philosophy: A Genealogy of Pragmatism*. London: MacMillan.

White, Alan R. 1972. "What We Believe." In *Studies in the Philosopy of Mind*. Ed. Nicholas Rescher. Oxford: Basil Blackwell, 1972, pp.69-84.

White, Morton. 1950. "The Analytic and the Synthetic: An Untenable Dualism." In *John Dewey Philosopher of Science and Freedom*. Ed. Sydney Hook. New York: The Dial Press, 1950, pp.316-330.

Wilson, Bryan R. ed. 1970. *Rationality*. New York: Harper Torchbooks.

Wilson, Catherine. 1992. "How Did the Dinosaurs Die Out? How Did the Poets Survive?" *Radical Philosophy* 62 Autumn: 20-24.

Wilson, Margaret. 1978. *Descartes*. London: Routledge & Kegan Paul.

Winch, Peter. 1958. *The Idea of a Social Science and Its Relation to Philosophy*. London: Routledge and Kegan Paul.

_____. 1970. "Understanding a Primitive Society." In Wilson (1970), pp.78-111.

Wisdom, John. 1957. *Philosophy and Psycho-Analysis*. Oxford: Basil Blackwell.

Wittgenstein, Ludwig. 1958. *Philosophical Investigations*. 2nd ed. Trans. G. E. M. Anscombe. Eds. G. E. M. Anscombe, R. Rhees and G. H. von Wright. Oxford: Basil Blackwell.

_____. 1972. *On Certainty*. Trans. Denis Paul and G. E. M. Anscombe. Eds. G. E. M. Anscombe and G. H. von Wright. New York: Harper Torchbooks.

_____. 1975. *Philosophical Remarks*. Trans. Raymond Hargreaves and Roger White. Ed. Rush Rhees. Chicago: University of Chicago Press.

_____. 1980. *Remarks on the Philosophy of Psychology, I*. Trans. G. E. M. Anscombe. Eds. G. E. M. Anscombe and G. H. von Wright. Chicago: University of Chicago Press.

_____. 1981. *Zettel*. 2nd ed. Trans. G. E. M. Anscombe. Eds. G. E. M. Anscombe and G. H. von Wright. Oxford: Basil Blackwell.

_____. 1983. *Remarks on the Foundations of Mathematics*. rev. ed. Trans. G. E. M. Anscombe. Eds. G. H. von Wright, R. Rhees and G. E. M. Anscombe. Cambridge, Ma.: MIT Press.

Woolf, Leonard. 1970. "Hunting the Highbrow." In *Hogarth Essays*. Eds. Leonard Woolf and Virginia Woolf. Freeport, New York: Books for Libraries Press.

Wright, Ronald. 1992. *Stolen Continents: The "New World" through Indian Eyes Since 1492*. Toronto: Viking.